Genetic improvement of grass pea (*Lathyrus sativus*) for low β-L-ODAP content

Peter Martin Ferdinand Emmrich

John Innes Centre

Thesis for the degree of Doctor of Philosophy (PhD)

Submitted for examination to the University of East Anglia in September 2016

Final version including minor corrections submitted in June 2017

Table of contents

Abstract

Grass pea (*Lathyrus sativus*) is a legume crop with great potential for global food security due to its exceptional tolerance to drought and flooding. The main limitation of this crop is the presence of the toxin β-L-oxalyl-2,3-diaminopropionic acid (β-L-ODAP) in its seeds and green tissues, which can cause paralysis in humans if grass pea is consumed over long periods. The objective of this study was to develop means to identify grass pea genotypes with reduced or zero seed β-L-ODAP content and to investigate the biosynthetic pathway of this compound in grass pea. To this end, collections of grass pea germplasm were screened for variation in seed β-L-ODAP levels. Considerable variation in β-L-ODAP levels was observed but no β-L-ODAP-free plants were identified. To increase the available variation for this trait, an EMS-mutagenised population was screened for low/zero-ODAP mutants. This mutant screen yielded 14 low-ODAP mutant lines, three of which were characterised using a mass spectrometry method, employing a stable-isotope-labelled isoform of β-L-ODAP as an internal standard. Both the development of the mass spectrometry method and the synthesis of the internal standard were performed for the purposes of this project. The three characterised lines yielded seed β-L-ODAP-contents below existing low-ODAP varieties, although none were β-L-ODAP-free. To further investigate the synthesis of β-L-ODAP, RNA was extracted from several tissues of grass pea and sequenced to create tissue specific transcriptomes. These were interrogated to identify candidate genes, which were tested using heterologous expression in *Nicotiana benthamiana*. One candidate gene of the BAHD-acyltransferase family was confirmed as an enzyme capable of catalysing the synthesis of β-L-ODAP. The identification of a set of low-ODAP mutants and the ODAP-synthase gene represent significant advances towards understanding the role of β-L-ODAP in grass pea and the development of grass pea genotypes free of this neurotoxin.

List of Abbreviations

AAE3	acyl-activating enzyme 3
AHCT	anthocyanin O-hydroxycinnamoyltransferase
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ATP	adenosyl triphosphate
BAHD acyltransferases	a superfamily of acyltransferases, the acronym is composed of the names of the first four enzymes of the family to be identified: BEAT, AHCT, HCBT, DAT
BAPN	β-aminopropionitrile
BARI	Bangladesh Agricultural Research Institute
BCKV	Bidhan Chandra Krishi Viswavidyalaya (Bidhan Chandra Agricultural University)
BEAT	benzylalcohol O-acetyltransferase
BIA	(β-isoxazolin-5-on-2-yl)-alanine
BLAST	Basic Local Alignment Sequence Tool
BOAA	β-oxalylaminoalanine (synonym of β-L-ODAP)
canola	CANada Low erucic Acid
CBNPMP	Conservatoire Botanique National des Pyrénées et de Midi Pyrénées
CLIMA	Centre for Legumes in Mediterranean Agriculture
CoA	Coenzyme A
CRISPR	clustered regularly interspersed short palindromic repeats
CZE	capillary zone electrophoresis
DABA	L-2,4-diaminobutyric acid
DAPA	L-2,3-diaminopropionic acid (synonym of L-DAP)
DAPRO	L-2,3-diaminopropionic acid (synonym of L-DAP)
DAT	deacetylvindoline 4-O-acetyltransferase
DNA	deoxyribonucleic acid

DZRC	Debre Zeit Research Centre
EBI	Ethiopian Biodiversity Institute
EDTA	ethylenediaminetetraacetic acid
EI	Earlham Institute (formerly The Genome Analysis Centre)
EIAR	Ethiopian Institute of Agricultural Research
EMS	ethyl methanesulphonate, a chemical mutagen
ENNG	N-ethyl-N-nitro-N-nitrosoguanidine, a chemical mutagen
FAO	Food and Agriculture Organisation
GABA	γ-aminobutyric acid
GCMS	gas chromatography mass spectrometry
GFP	green fluorescent protein
GMO	genetically modified organism
GPC	Global Pulse Confederation
HCBT	anthranilate N-hydroxycinnamoyl/benzoyltransferase
HPLC	high-performance liquid chromatography
HQI	hydrargyrum quartz iodide
HTML	hypertext markup language
IAEA	International Atomic Energy Agency
ICARDA	International Centre for Agricultural Research in the Dry Areas
INRA	Institut National de la Recherche Agronomique
IPK	Institut für Pflanzenkunde
JIC	John Innes Centre
LB	lysogeny broth
LCMS	liquid chromatography mass spectrometry
L-DAP	L-2,3-diaminopropionic acid
MES	2-(N-morpholino)ethanesulfonic acid
NCBI	National Center for Biotechnology Information
NIAB	National Institute for Agricultural Botany
NMR	nuclear magnetic resonance
NMU	N-nitroso-N-methylurea
OAP	3-oxalylamino-2-aminopropionic acid (synonym of β-L-ODAP)
OAS	O-acetylserine
ODAP	L-N-oxalyl-2,3-diaminopropionic acid, present as α- and, β-isomers

OPA	o-phthalaldehyde
ORF	open reading frame
OsDREB	*Oryza sativa* dehydration responsive element binding gene
OTU	operational taxonomic unit
PCR	polymerase chain reaction
pEAQ-HT	pEAQ-Hypertrans, a system of plasmid vectors for high efficiency transient expression
PGS	International Consortium for Pea Genome Sequencing
PTT	potassium tetraborate tetrahydrate
PVP	polyvinylpyrrolidone
QTL	quantitative trait locus
RNA	ribonucleic acid
RO	reverse osmosis
SCPL acyltransferases	Serine carboxypeptidase-like acyltransferases
SDS	sodium dodecylsulfate
SIL	stable isotope labelled
TALEN	transcription-activator-like-effector nuclease
TBE	Tris/borate/EDTA
TGAC	The Genome Analysis Centre (now the Earlham Institute)
TILLING	targeting induced local lesions in genomes
TMS	trimethylsilyl
UN	United Nations
USDA	United States Department of Agriculture
WGS	Whole Genome Shotgun sequencing
β-L-MDAP	β-L-malonyl-diaminopropionic acid

Acknowledgements

List of figures

List of tables

Chapter 1 – General introduction

1.1 Food security in the 21st century

1.1.1 The need for increased diversity of food production

About 2500 species of crop plants have undergone domestication (Meyer et al., 2012), but the vast majority of global food and feed production is covered by just a small number of crop species. Two thirds of global food calorie production involves just four crop species: maize, wheat, rice and soybean (Tilman et al., 2011; Ebert, 2014). Major crop species such as these have undergone extensive breeding in traditional farming systems and more recently in scientifically-directed breeding programmes. These efforts have led to vast improvements in crops, particularly with regard to yield, disease resistance and geographic range. The advances made during the green revolution have allowed global agricultural systems to outpace the growth in demand for food despite the fourfold increase in the human population over the course of the twentieth century (Krausmann et al., 2009). However, during this period, global food supplies have become more homogenous, with increasing reliance on a small number of species supplying the majority of calories for human nutrition (Khoury et al., 2014).

To keep up with the rise in global human and livestock populations, global food and feed production must be increased by 100-110% between 2005 and 2050 (Tilman et al., 2011), a goal we are unlikely to reach, according to current trends in crop yields (Ray et al., 2013). Currently, a quarter of global agricultural land suffers from water stress, a figure that rises to 40% when considering only land under irrigation (Gassert, 2013). The amount of land suffering water stress through periodic or chronic drought is likely to increase over the course of the 21st century due to the effects of climate change (Dai, 2013). This means that crops that are able to deliver high yields under conditions of water stress will be essential to maintain and increase yields as climate change progresses. This has made improved drought tolerance a major objective in crop development by breeding (Cattivelli et al., 2008; Langridge and Reynolds, 2015; Tuberosa, 2012) and genetic engineering (Hu and Xiong, 2014).

1.1.2   The potential of *Lathyrus sativus*

A complementary approach to the continued improvement of major crop species to maintain sustainable food and feed production in the face of climate change is to increase the use of currently underutilised crops that already show favourable characteristics in terms of tolerance to drought and other environmental stress factors (Massawe et al., 2016). One such species is *Lathyrus sativus,* commonly called grass pea, a grain legume species and the most commonly cultivated species of the diverse genus *Lathyrus,* which includes ~160 other species (Asmussen and Liston, 1998). Other names for this crop include khesari (India and Bangladesh), guaya (Ethiopia), 家山藜豆 jia shan li dou (China), cicerchia (Italy), almorta (Spain) and chickling vetch (UK and USA). Grass pea is currently grown on an estimated 1.50 million hectares worldwide, with an annual production of 1.20 million tonnes (Kumar et al., 2011a), though as most of this production occurs on marginal land owned by poor smallholder farmers, accurate figures are difficult to gather. Most of the production is in South Asia, in particular Bangladesh, India, Nepal and Afghanistan as well as Ethiopia and Eritrea. Reported yields under experimental conditions vary widely from 420 to 3860 kg/ha/yr in Italy (Piergiovanni et al., 2011); from 1079 to 1583 kg/ha/yr in Turkey (Karadag and Yavuz, 2010); from 1318 to 1795 kg/ha/yr in India (Campbell, 1997) and up to more than 5 t/ha/yr in variety tests under favourable conditions in Ethiopia (Tsegaye et al., 2005). This indicates that there is a substantial gap between the yields commonly achieved in existing farming systems and the potential yields that can be achieved by using improved varieties and higher farming inputs. More informative than these yield figures under experimental conditions are yields achieved in realistic farm settings. In the common 'utera' farming setting, where grass pea is sown by broadcasting seed into a standing rice crop and grows up between the rice straw, yields between 509 and 783 kg/ha/yr have been observed. Across all farm settings in the Indian state of West Bengal, average yields in the period of 2001-2011 have ranged from 628 to 1172 kg/ha/yr (personal communication from Prof S.K. Samanta, Bidhan Chandra Agricultural University (BCKV), West Bengal).

Beyond its yield potential, grass pea has been claimed as having remarkable yield stability compared to other legume crops (Zhelyazkova et al., 2016) in the face of drought and flooding. It has also been noted anecdotally for its tolerance to salinity, heat stress, insect herbivory and some fungal diseases (Yang and Zhang, 2005; Vaz Patto et al., 2006), although detailed comparative studies involving several species in multiple environments are limited. *Lathyrus sativus* appears to be better able to maintain photochemical efficiency

during mild and severe water stress than other legume crops (Silvestre et al., 2014). These characteristics have allowed grass pea to be cultivated in agricultural settings unsuitable for other crops, especially on marginal cropland with poor soils and unpredictable rainfall patterns. Today, grass pea is an important food crop in smallholder-agriculture in India, Bangladesh, Nepal, Ethiopia and Eritrea and is also cultivated in China, Pakistan, Afghanistan, Australia, the Mediterranean circumference, Eastern Europe and Chile as both a feed and food crop (Campbell, 1997). It is capable of highly efficient nitrogen fixation through its symbiosis with *Rhizobium leguminosarum* bv. *viciae* (Jiao et al., 2011a; Drouin et al., 2000). A nitrogen fixation rate of 124 kg/ha/yr has been recorded (Schulz et al., 1999), making grass pea a highly efficient crop for increasing soil nitrogen, especially in dry conditions. Grass pea can be grown in low-input agriculture in marginal areas and makes a crucial contribution to the food security of poor smallholder farmers (Girma et al., 2011; Haque et al., 1996; Bhowmick, 2013). Its use as an insurance crop for times when natural disasters wipe out all other food sources has made grass pea a life-saver for millions of people during times of food shortages (Girma et al., 2011; Girma and Korbu, 2012). Grass pea is an important crop during lean times in Bangladesh, as it can be harvested as a nutritious leafy vegetable during the lean season when few other foods are available (Hussain, 1989). Grass pea produces seed with very high protein concentration of up to 29.9 % w/w as measured by the Kjeldahl method (Bisignano et al., 2002; Aletor et al., 1994) and is rich in lysine. The quality of the grass pea protein is limited only by its low levels of sulphur-containing amino acids (methionine and cysteine) as well as tryptophan (Pastor-Cavada et al., 2011), as with other grain legumes (Iqbal et al., 2006). The spectrum of essential amino acids in grass pea complements the spectrum found in most cereals, which are rich in sulphur-containing amino acids, but poor in lysine. Grass pea is an excellent low-input crop: it can be sown by broadcasting; it does not require irrigation or fertilizer and is less susceptible to insect pests than most other legumes (Campbell, 1997). This makes it ideal for smallholder farmers who cannot utilize the economies of scale necessary to grow high-input crops profitably and who do not have the necessary education to correctly apply complicated agricultural management practises. Grass pea seeds contain significant quantities of the non-protein amino acid homoarginine, which is catabolised into the vasodilating hormone nitric oxide and is associated with cardiovascular health benefits (Rao, 2011).

1.2 History of grass pea cultivation

1.2.1   From Neolithic staple to orphan crop

Grass pea is an ancient crop species. Based on archeobotanical and phytogeographical evidence, the origin of cultivation of *L. sativus* has been suggested to be the Balkan peninsula (Kislev, 1989). No truly wild *Lathyrus sativus* persists today, although the plant occasionally appears as a weed among other crops and growing in the wild as escapees from cultivation (Jackson and Yunus, 1984). The species *L. sativus* is likely derived from its closest wild relative *L. cicera* (red pea) (Belaid et al., 2006). Fossilised seeds of *L. sativus* and allied species (*L. cicera* and *L. ochrus*) have been found in several Neolithic sites in the fertile crescent, Eastern Europe and the Mediterranean, dated to earlier than 6000 BCE (Coward et al., 2008; Kislev, 1989; Peña-Chocarro et al., 2013; Marinova, 2007). The earliest fossils of grass pea used for human consumption date back to the middle pre-pottery Neolithic era (8200-7500 BCE) in the Euphrates valley (Ferrio et al., 2012). This places grass pea among the earliest crops to be domesticated by humans. Evidence of grass pea cultivation has been found at bronze age sites in India, Greece (Valamoti et al., 2010), the Levant (Van Zeist and Bakker-Heeres, 1985; Mahler-Slasky and Kislev, 2010) and Ethiopia (Butler et al., 1999). Despite this ancient origin of the crop, grass pea is not currently grown as a major crop globally and the acreage under cultivation with grass pea has been decreasing. Comparatively little scientific research or breeding efforts have been focused on it, and hence genetic resources remain very limited. Several sets of genetic markers have been described for grass pea (Chapman, 2015; Lioi and Galasso, 2013; Shiferaw, 2013; Sun et al., 2012; Yang et al., 2014), but no detailed genetic map has been released to date. Limited transcriptomic data are available (Chapman, 2015; Matasci et al., 2014), but do not cover all tissues of the plant and no genome sequence for grass pea is available, to date. Modern grass pea cultivars retain many unwanted characteristics that have been bred out of other legume crops such as pea. These include indeterminate growth, sprawling growth habit, single podding and a low harvest index (Campbell, 1997) as well as high contents of trypsin inhibitors (Wang et al., 1998) and tannins (Aletor et al., 1994). One hypothesis proposed to explain this lack of success in improvement of grass pea is that the conflicting breeding goals of improving grass pea as a seed crop (by improving seed yield) and as a forage crop (by improving herbage yield) have cancelled each other out, leading to little substantial improvement in either direction (Smartt, 1984).

Figure 1. Painting of a flowering shoot and developing pod of Lathyrus sativus. Reproduced from Curtis's Botanical Magazine vol. 3-4 1790/91

1.2.2   Lathyrism – a neurodegenerative disease caused by grass pea

The primary issue limiting the use of grass pea as a food security crop is its association with a debilitating neurodegenerative disease, which can occur if grass pea seeds comprise more than a third of the caloric intake of a human for at least three months (Dufour, 2011). This disease has been observed since antiquity, with the earliest surviving description by the Greek physician Hippocrates (460-377 BCE), followed by other ancient sources including the Greek physician Galen, the Roman naturalist Pliny the Elder and the Persian polymath Ibn Sīnā (lat. Avicenna) (Dastur and Iyer, 1959; Hendley, 1903). The disease is marked by an irreversible spastic paraparesis of the legs, followed by muscular atrophy (Khan et al., 1995). Patients present with a scissor-like gait and further progression of the disease leads to complete immobilisation of the legs. Mild symptoms can be reversed if discovered early and grass pea consumption is stopped, but once the disease has progressed, the condition becomes permanent. Recent neuroimaging data suggest that the aetiology of the disease is due to spinal motor neurons rather than the motor cortex in the brain (Meiner and Gotkine). As the disease is much more likely to affect people suffering from malnourishment, particularly with regard to sulphur-containing amino acids (Van Moorhem et al., 2011), this can have a devastating impact among the poorest, who often rely heavily on grass pea in countries such as Ethiopia, India, Bangladesh and Afghanistan (Dufour, 2011; El-Moneim et al., 1999).

This disease was named 'lathyrism' by the Italian physician Arnaldo Cantani in 1873 (Cantani, 1873). The diagnosis is now commonly specified as neurolathyrism, to distinguish it from the separate syndromes osteolathyrism and angiolathyrism, both caused by a deficiency in collagen crosslinking, primarily induced by the toxin β-aminopropionitrile (BAPN) (Dasler and Stoner, 1959; Lees et al., 1990). BAPN is produced by the catabolism of 2-cyanoethyl-isoxazolin-5-one, a compound which is found in green tissues of several species of the *Lathyrus* genus, especially *Lathyrus odoratus* (Ikegami et al., 1984). Symptoms of osteolathyrism include a weakness of connective tissues, resulting in bone deformity. When affecting the cardiovascular system (angiolathyrism), the reduced collagen crosslinking can result in aortic aneurisms. Other compounds, which are not present in *Lathyrus* tissues, including semicarbazide, thiosemicarbazide, benzoic hydrazide and aminoacetonitrile, have also been identified as osteolathyrogens (Dawson et al., 2002; Dawson et al., 1990; Chowdhury and Davis, 1989; Schultz and Ranney, 1988). Symptoms of osteolathyrism have been observed among a minority of patients suffering neurolathyrism in Bangladesh (Haque et al., 1997). However, the levels of 2-cyanoethyl-isoxazolin-5-one in

Lathyrus sativus are low and the disease is more commonly associated with the consumption of *Lathyrus odoratus*, hence its alternative name 'odoratism' (Dastur and Iyer, 1959). The concentrations of 2-cyanoethyl-isoxazolin-5-one in grass pea tissues were not studied as part of this thesis.

The risk of neurolathyrism has led to grass pea being considered a food of last resort and a food only for the poor in many countries where it is cultivated. To contain the risk of neurolathyrism, several governments have legislated against the cultivation or sale of grass pea. The first such ban was implemented by an edict by Georg the Duke of Wurttemberg in 1671, which had to be reinforced by Georg's successors, as the use of grass pea in the country continued (Cohn and Streifler, 1983). By the beginning of the 20th century, several other countries had placed restrictions on the cultivation or the sale of grass pea (Hendley, 1903). Grass pea used to be a traditional crop in several provinces of China. During the drought in central China in the 1970s, grass pea was used heavily to sustain farming populations, but this was followed by an epidemic of lathyrism that led to the crop being outlawed. Recent interest in grass pea as a crop for agriculture in marginal soils, particularly for feed uses, has brought about calls for the reintroduction of grass pea to China (Yang and Zhang, 2005). After several epidemics of neurolathyrism following periods of drought, the governments of some Indian states banned the sale of grass pea, as well as the use of grass pea to pay the wages of workers. The production of the crop in subsistence agriculture was not banned and continued, as did the (outlawed) use of grass pea flour as an adulterant for more expensive besan (chickpea) flour (Pradesh, 2008). Due to improvements in infrastructure and farm yields, the diversity of food sources available to the poorest in India has improved over the last thirty years. As the number of people relying solely on grass pea has decreased, so has the number of cases of lathyrism. This has led some observers to consider lathyrism a threat of the past. Yet, epidemics of neurolathyrism have occurred during times of extreme drought when other crops have become unavailable, such as during the 1998/99 famines in Ethiopia (Woldeamanuel et al., 2012; Getahun et al., 1999; Haimanot et al., 2005) and Afghanistan (Simpson, 2002).

The disease disproportionately affects young men (Hendley, 1903; Haque et al., 1996; Dwivedi and Prasad, 1964; Haimanot et al., 2005; Haimanot et al., 1990), but the reasons for this gender disequilibrium are unclear. In most cases of neurolathyrism among farming populations, only a small proportion (~6%) of the population that relies heavily on grass pea develops neurolathyrism (Getahun et al., 1999; Haque et al., 1996; Dufour, 2011). In a comprehensive study involving more than a million people in northern and central Ethiopia

(the primary areas of grass pea cultivation in the country), a neurolathyrism prevalence of
0.3 % was observed, ranging from 0.01 % to 0.75 % between districts (Haimanot et al.,
1993). This means that in conditions where no other food sources are physically accessible
or affordable by the poor, the risk of neurolathyrism may be worth taking if the alternative
is starvation. The two-edged nature of grass pea as both life-saving crop during times of
famine and war, but also a toxic crop that can gravely endanger people who consume it, is
exemplified in the engraving *Gracias á la almorta* (Thanks to the Grass pea) by Francisco de
Goya, which is reproduced in Figure 2. The engraving shows Spanish civilians eating grass
pea, with one of them already suffering from paralysis. During a famine in Madrid in 1811-
12, which occurred during the Peninsular war (1807-1814), grass pea remained one of the
few foods still available.

Figure 2. Gracias á la almorta - engraving by Francisco de Goya, part of the series Los Desastres de la Guerra. Depicted are Spanish civilians sharing a meal of grass pea during a famine in 1811/12. The woman in the foreground appears to suffer symptoms of neurolathyrism.

1.3   The neurotoxin β-L-ODAP

The compound β-N-oxalyl-L-α,β-diaminopropionic acid (β-L-ODAP), (structure shown in Figure 3) has been identified as the causative agent of neurolathyrism (Rao et al., 1964; Murti et al., 1964; Spencer et al., 1986). In medical literature, β-L-ODAP is often referred to as β-N-oxalyl-amino-alanine (BOAA) (Aletor et al., 1994; Gannon and Terrian, 1989; Ormandy and Jope, 1990; Pai and Ravindranath, 1993; Ross et al., 1989; Weiss et al., 1989) and occasionally as L-3-oxalylamino-2-amino propionic acid (OAP) (Mehta et al., 1983). The compound induced acute convulsions in 1-day-old chicks (Rao et al., 1964) and hind-leg paralysis in squirrel monkeys (*Saimiri sciureus*) when introduced into the cerebrospinal fluid by lumbar puncture (Lakshmanan et al., 1971). No symptoms were observed following chronic oral administration of up to 6 mg/g body weight/day β-L-ODAP to squirrel monkeys, although an increase in the dose to 8 mg/g/day led to acute neurotoxicity causing the death of experimental animals after 3-5 days. This study was based on a small sample size of only three animals (Mehta et al., 1983). The ethical implications of testing on primates make this species unsuitable as an animal model. Symptoms akin to human neurolathyrism (spastic paraparesis of the legs) have been induced in rats though repeated injection of small doses of β-L-ODAP into the skin at the back of the animal (Kusama-Eguchi et al., 2005; 2010).

Neurotoxins

A neurotoxin is a substance that is poisonous or destructive to nerve tissues. Neurotoxic exposures that threaten human health can be the result of a wide range of factors, including: air pollution (e.g. ultrafine particulate matters, ozone); toxic heavy metals in drinking water (e.g. as a result of soil acidification or industrial runoffs); medical or recreational drug use (e.g. opioid abuse); venomous exposures (e.g. snake bites) and dietary uptake (including plant-derived neurotoxins such as cyanogens from cassava or β-L-ODAP from grass pea, microbial products such as ethanol or botulinum toxin and animal-derived neurotoxins such as pufferfish tetradotoxin) (Tshala-Katumbay and Spencer, 2007; 2015).
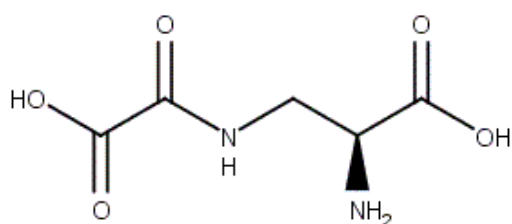
Figure 3. Structural formula of β-L-ODAP

β-L-ODAP appears to act as an analogue to glutamate, which serves as a common neurotransmitter in animals. This causes β-L-ODAP to interfere with the transport of glutamate and its decarboxylation into γ-aminobutyric acid (GABA) (Ross et al., 1985), which may explain the excitotoxic (convulsive) effects observed in some experimental animals, although these symptoms are not apparent in humans. β-L-ODAP also acts as a competitive antagonist to glutamate on α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, causing neurotoxic effects in rat brains (Ross et al., 1989; Kusama-Eguchi et al., 2010). An isomer of β-L-ODAP, α-L-ODAP is also found in grass pea tissues, but makes up only around 5 % w/w of total ODAP (Arentoft and Greirson, 1995; Roy and Rao, 1968). The α-isomer does not appear to be neurotoxic (Chase et al., 1985). Besides several species of the *Lathyrus* genus, β-L-ODAP is found in species of the legume genera *Acacia* and *Crotalaria* (Quereshi et al., 1977; Evans and Bell, 1979), as well as the unrelated genus *Panax* (Kuo et al., 2003).

One of the most striking descriptions of the effects of β-L-ODAP consumption on human physiology stems from one of the darkest chapters of history. Between 1941 and 1944, several thousand Romanian and Ukrainian Jews were held at the concentration camp at Vapniarka, Ukraine, by the fascist government of Romania. During this period, prisoners were fed from a leftover consignment of horse-feed, resulting in a highly deficient diet, consisting primarily of 160 g of barley and 200 g of boiled grass pea per day with an estimated β-L-ODAP-concentration of 0.25 % w/w (Lambein et al., 2001). The grass pea ration was later increased to 400 g per day. Within one month of the ration being increased, the first of the prisoners started experiencing symptoms of neurolathyrism. Over the following six months, an epidemic of neurolathyrism developed, with up to 60 % of the prisoner population showing symptoms – a much higher proportion than normally

observed among human populations heavily dependent on grass pea (Getahun et al., 1999; Haimanot et al., 2005; Dwivedi and Prasad, 1964; Haque et al., 1996). The increased ration corresponded to an estimated dose of 1 g β-L-ODAP/day (Lambein et al., 2001). The physician Dr Arthur Kessler, himself one of the prisoners, diagnosed and documented the spread of the disease (Westmore and Weisz, 2013). He noted that the incidence of neurolathyrism was lower among the prisoners who were made to work in the forest outside the camp and conjectured that the small amounts of roots, berries and tree nuts that these prisoners ate while working helped to partially protect them against neurolathyrism (Lambein et al., 2001). This protective effect may be due to the presence of sulfur-containing amino acids, which present at very low concentrations in grass pea but have been shown to protect against β-L-ODAP toxicity in cell culture experiments (Kusama-Eguchi et al., 2011), in these foraged foods. The data collected by Dr Kessler during his internment provide the best estimate of the toxicity of β-L-ODAP to humans and underscore the importance of malnutrition in the aetiology of neurolathyrism.

Another toxin, L-2,4-diaminobutyric acid (DABA), which also induces symptoms of neurolathyrism is found in several *Lathyrus* species, in particular *L. sylvestris* (Rowe et al., 1993), *L. latifolius* (Barrow et al., 1974) and *L. hirsutus* (Holbrook et al., 2015) but not in *L. sativus* or *L. cicera* (Bell, 1964). DABA toxicity is a problem for grazing animals feeding on wild *Lathyrus* species (Holbrook et al., 2015), but very rarely affects humans, as these species are not used as food crops.

1.3.1   Biosynthesis of β-L-ODAP in grass pea

Since the identification of β-L-ODAP as the toxin responsible for neurolathyrism, several research groups have investigated the biosynthetic pathway by which this compound is produced in grass pea, primarily using enzyme extracts (Malathi et al., 1968; Malathi et al., 1970; Malathi et al., 1967; Ikegami et al., 1993) and radioisotope feeding (Kuo et al., 1994; Kuo et al., 1998; Kuo and Lambein, 1991; Lambein et al., 1990). This research led to the pathway shown in Figure 4 being proposed. Not all steps in the pathway have been proven experimentally and some of the intermediates remain hypothetical. The pathway appears to branch from primary metabolism with the formation of β-(isoxazolin-5-on-2-yl)-alanine (BIA) from O-acetyl-serine and the (hypothetical) intermediate isoxazolin-5-one, a reaction catalysed by a cysteine synthase (Ikegami et al., 1991). BIA is also found in many other legumes, including pea (*Pisum sativum*) (Schenk and Werner, 1991) and lentil (*Lens culinaris*) (Kuo et al., 1998). Carbon-14 introduced in the form of radiolabelled BIA is

incorporated into β-L-ODAP in callus tissue and developing pods of grass pea (Kuo and Lambein, 1991; Kuo et al., 1994; Lambein et al., 1990), but the exact steps of the biosynthesis are not known. The formation of L-2,3-diaminopropionic acid (L-DAP, also referred to as DAPRO or DAPA) has been demonstrated in vitro using an enzyme extract from grass pea (Ikegami et al., 1999), but attempts to prove the presence of this compound in grass pea tissues have been unsuccessful. It is believed that this synthesis intermediate is short-lived and does not accumulate to detectable levels in grass pea. As early as the late 1960s, Malathi et al. were able to partially purify two enzymes from grass pea that catalysed the ATP-dependent synthesis of oxalyl-coenzyme A from oxalic acid and coenzyme A and the formation of β-L-ODAP from L-DAP and oxalyl-CoA during in vitro experiments. These enzymes were labelled oxalyl-CoA synthetase and β-L-ODAP-synthase, respectively (Malathi et al., 1968; 1970; 1967). However, the genes encoding these enzymes have not been identified or located on a genetic map of grass pea to date. If these enzymes can be extracted again and sufficiently purified it may be possible to ascertain their amino acid sequence and use it to identify the genes encoding these enzymes in grass pea.
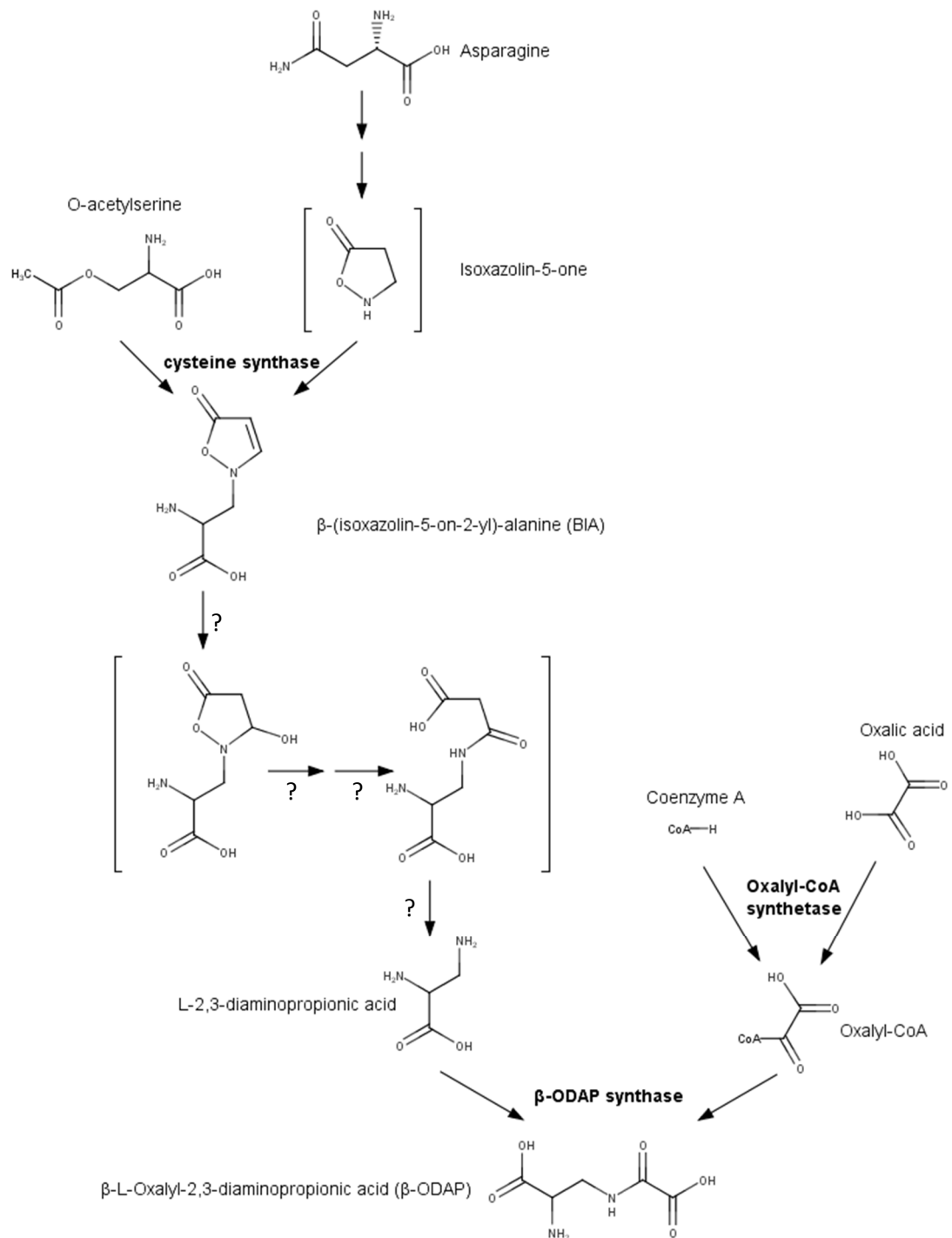
Figure 4. Proposed pathway for β-L-ODAP synthesis. Redrawn based on publications by Yan et al., Malathi et al., Kuo, Ikegami and Lambein (Malathi et al., 1970; Ikegami et al., 1991; Ikegami et al., 1999; Kuo et al., 1998; Kuo et al., 1994; Kuo and Lambein, 1991; Yan et al., 2006). Hypothetical intermediates are shown in square brackets.

1.3.2   The ecophysiological role of β-L-ODAP

Several hypotheses have been proposed for the role of β-L-ODAP in the physiology of grass pea and its interaction with the environment, but evidence for these remains slight. The most obvious possibility is that β-L-ODAP serves as a defence compound against shoot herbivores. This is supported by the pattern of β-L-ODAP concentrations observed in seedlings and tissues of juvenile and mature plants. The toxin accumulates primarily in young tissues, which are vulnerable to herbivores but of importance to the plant's further potential photosynthetic activity (Jiao et al., 2006; Kuo et al., 1994; Xiong et al., 2015). In most mature tissues, toxin levels are very low. This mirrors the distribution of defence compounds in other species (Ballhorn et al., 2009). Purified β-L-ODAP has been shown to reduce the growth of insect larvae of the rice moth (*Corcyra cephalonica)* (Rao et al., 1964). Yet to date, no study has been published confirming β-L-ODAP as an insect feeding inhibitor in *L. sativus* although a weak effect of inhibited feeding due to β-L-ODAP and other non-protein amino acids has been observed in *Acacia* ssp. and *L. latifolius* (Bell et al., 1996). When exposed to storage pests, reduced-ODAP varieties of grass pea did not show increased infestation (Roy and Bhat, 1975)

The inclusion of large amounts of grass pea in the diets of livestock leads to negative effects on growth (Hanbury et al., 2000; Enneking, 2011), but it is unclear whether this is due to β-L-ODAP or other antinutritional factors (Aletor et al., 1994; Roy and Bhat, 1975; Wang et al., 1998). It is unclear whether the diet of any large herbivore in the wild would be so heavily reliant on grass pea that the relatively low toxicity of β-L-ODAP would result in an evolutionary advantage to the plant.

Another proposed biological function is that β-L-ODAP is involved in responses to oxidative stress, as evidenced by the strong negative correlation between β-L-ODAP levels and reactive oxygen species in leaves (Jiao et al., 2011b). This could mean that β-L-ODAP is directly connected to the remarkable drought tolerance of grass pea. However, other research did not find any direct relationship between β-L-ODAP and free radical metabolism (Xing et al., 2001). As β-L-ODAP can accumulate to high concentrations in plant tissues without becoming toxic to the plant (Jiang et al., 2013; Jiao et al., 2006), it is possible that the compound acts as an osmoprotectant to prevent osmotic changes causing cell damage. One study found increased levels of reactive oxygen species scavenging molecules and other osmoprotectants in grass pea compared to pea in response to drought (Jiang et al., 2013).

β-L-ODAP acts as a chelator of bivalent metal cations such as Zn^{2+}, Cu^{2+}, Fe^{2+} and Mn^{2+}. In addition, the levels of β-L-ODAP in grass pea tissues show a marked response to the availability of certain metal ions in the soil (Lambein et al., 1994; Xiong et al., 2014; Lambein, 2000). The chelating action of β-L-ODAP may be related to its toxicity in animals (Lambein et al., 1994).

Because the biological role of β-L-ODAP in grass pea remains controversial, it is unclear what effect a mutation that completely disrupts the production of this compound would have on the physiology and agricultural utility of the mutant plant. It is conceivable that grass pea plants without β-L-ODAP would be highly susceptible to certain pests or would be compromised in their ability to withstand abiotic stresses. However, no increased susceptibility to biotic or abiotic stress factors has been recorded for previously developed low-ODAP varieties. Genotypes with 80-90 % lower β-L-ODAP concentrations in their tissues than common landraces are now being cultivated successfully in several countries (Akter et al., 2015; Kumar et al., 2011a; Siddique et al., 2006). The most direct route to investigate the ecophysiological role of β-L-ODAP in grass pea remains to develop β-L-ODAP-free genotypes and test their performance under various biotic and abiotic stress conditions alongside near-isogenic genotypes containing β-L-ODAP.

1.3.3   Mitigation of grass pea toxicity through food processing

The toxin β-L-ODAP is contained in all organs of the plant, with highest concentrations in young shoot, leaf and developing pods (Jiao et al., 2006). Grass pea seeds are commonly consumed in a variety of forms, such as dhal, sauce (e.g. the Ethiopian shiro wot), or as roasted and salted immature pods eaten as a snack (Dufour, 2011). To a lesser extent, grass pea shoots are eaten as a leafy vegetable, usually after steaming or boiling. These methods of preparation reduce the toxicity to some extent by leaching out the toxin through steeping or boiling and discarding the water or by partial conversion of β-L-ODAP into nontoxic α-L-ODAP (Padmajaprasad et al., 1997). Boiling and fermentation with the fungal species *Rhizopus oligosporus* and *Aspergillus oryzae* appear to be more effective at reducing toxicity than roasting, but do not result in complete detoxification (Padmajaprasad et al., 1997; Ramachandran et al., 2005; Yigzaw et al., 2004; Kuo et al., 1995). Ethnographic studies in Ethiopia and India show that, besides some inaccurate beliefs (for example that the steam coming off a pot of boiling grass peas is highly toxic), most consumers are well aware of the dangers associated with eating grass pea and use specific preparation techniques to reduce toxicity (Girma et al., 2011; Butler et al., 1999).

However, no method of food preparation is able to detoxify grass pea entirely. In addition, consumers may be forced to switch to food preparation methods that are less efficient at reducing toxicity (e.g. by not discarding the water after boiling grass pea seeds) if water or fuel are limited, as is common during times of drought.

1.3.4   Physiological breakdown of β-L-ODAP

There appears to be a catabolic pathway for the breakdown of β-L-ODAP in humans, which is less efficient or non-existent in other animals. Human volunteers who consumed cooked grass pea seeds or controlled amounts of pure β-L-ODAP excreted less than 1 % of it in their urine, but increased their excretion of oxalic acid (Pratap Rudra et al., 2004). In mice, rats and chicks, much greater proportions (between 21.1 % and 75.2 %) of orally or intraperitoneally administered radiolabelled β-L-ODAP were excreted in urine (Jyothi et al., 1998). The breakdown of β-L-ODAP in humans may provide an explanation for the typically observed low incidence of neurolathyrism in human populations relying heavily on grass pea (Barrow et al., 1974; Dufour, 2011; Girma et al., 2011; Haque et al., 1996). It is also possible that there is genetic variation in the susceptibility to neurolathyrism among the human population (Pratap Rudra et al., 2004), but no study to identify genetic markers associated with susceptibility to neurolathyrism has been conducted.

1.3.5   Potential biomedical applications of β-L-ODAP

β-L-ODAP may have bioactive properties that could make it a useful compound for biomedical applications. In addition to species of the genera *Lathyrus*, *Acacia* and *Crotalaria*, β-L-ODAP is also found as an abundant free amino acid in ginseng (*Panax spp.*) where it is commonly referred to by the trivial name dencichine (Kuo et al., 2003). This has led to investigations of whether low doses of β-L-ODAP could be associated with the health benefits ascribed to ginseng (Lambein, 2000). β-L-ODAP has known hemostatic properties, which appear to be due to a vasoconstrictive effect, rather than any effect on blood clotting (Okuda et al., 1990). To capitalise on this bioactive property, the Chinese company Yunnan Baiyao (Kunming, Yunnan Province) is developing band-aids containing β-L-ODAP to reduce blood flow (Wang et al., 2014). In addition, β-L-ODAP is being investigated for bioactive properties that may be beneficial in the treatment of hypoxia (Eslavath et al., 2016) and Alzheimer's disease (Rao, 2011) and has recently been patented as a drug for the treatment of thrombocytopenia (Lan et al., 2016).

1.4 Low-ODAP genotypes of grass pea

1.4.1 Released low-ODAP varieties of grass pea

There is considerable variation for β-L-ODAP concentrations among grass pea genotypes. Seed β-L-ODAP concentrations ranging from 0.1% to 2.5 % w/w have been reported (Yadav and Mchta, 1995; Kumar et al., 2011a). The most commonly applied threshold for low-ODAP genotypes is 0.1 % of dry seed weight (Asthana, 1995; Chakrabarti et al., 1999) and several varieties that are reported to have β-L-ODAP concentrations below this threshold have now been released (Siddique et al., 2006). These include the low-ODAP varieties BioL-212 (Ratan), Mahateora, LS 8246 and Prateek, which were derived from Indian germplasm through breeding and (in the case of Ratan) selection of somaclonal variation in tissue culture (Kumar et al., 2011a; Sawant et al., 2011; Chakrabarti et al., 1999; Santha and Mehta, 2001; Tsegaye et al., 2005). Somaclonal variation is genetic variation caused by the mutagenic effect of tissue culture which may result, for example, in chromosome rearrangements or transposable element activations (Bairu et al., 2011). A new low-toxin variety, Bidhan Khesari 1, is currently under development at Bidhan Chandra Agricultural University (BCKV) in the state of West Bengal (personal communication, Abhimanyu Sarkar, JIC). Low-toxin varieties have also been released by the Bangladesh Agricultural Research Institute (BARI) under the names BARI Khesari 1, 2 and 3 (Akter et al., 2015). One low-toxin variety, Ceora, was developed by the Centre for Legumes in Mediterranean Agriculture (CLIMA) in Australia (Siddique et al., 2006). Although grass pea is cultivated to a considerable extent in Ethiopia, only one low-ODAP variety of grass pea (Wasie) has been released in this country to date (Kumar et al., 2011a). This variety was produced through a collaboration between the Ethiopian Institute for Agricultural Research (EIAR) and the International Centre for Agricultural Research in the Dry Areas (ICARDA). Several accessions are currently undergoing multi-location trials conducted by EIAR, with a focus on improving agronomic characteristics and stability in β-L-ODAP content to enable the development of future improved low-ODAP varieties (personal communication, Alemu Abate, Aksum University, Ethiopia).

These varieties represent an important advance in the improvement of grass pea. Their low β-L-ODAP levels reduce the risk of neurolathyrism among consumers. However, none of these varieties have achieved the goal of an entirely β-L-ODAP-free crop. A strong genotype x environment interaction has been observed for the production of β-L-ODAP. This causes

β-L-ODAP levels to be increased significantly under conditions of stress, in particular cultivation in soils low in Zn^{2+} or high in Fe^{3+} as well as under conditions of drought (Fikre et al., 2006; Fikre et al., 2011; Polignano et al., 2009). This has raised justified concern about whether these lines will be safe to consume when they are grown by smallholder farmers under extreme environmental conditions (Fikre et al., 2008).

1.4.2   Attempts to produce ODAP-free genotypes

Since the discovery of the toxin by Rao et al. (1964) all attempts to find β-L-ODAP-free grass pea genotypes by screening natural germplasm, have been unsuccessful (Kumar et al., 2011a). This makes studies of the function of β-L-ODAP and its role in the aetiology of lathyrism after prolonged grass pea consumption very difficult. A few mutant screens have been performed with grass pea, but none of them have focused on finding a low-/zero-β-L-ODAP line (Rybiński, 2003; Talukdar, 2009b) or they have screened only very small numbers of mutant families without identifying stable low-ODAP lines (Nerkar, 1972, 1973, 1976). The main difficulties are the resources needed to grow a large population of mutants up to maturity and the lack of high-throughput screening methods.

Several research groups are now hoping to develop transgenic grass pea lines that are free of β-L-ODAP, by adding genes encoding enzymes that degrade β-L-ODAP or one of the intermediates of its synthesis (Kumar et al., 2016; Yadav and Mchta, 1995). Another option for reducing β-L-ODAP production by a transgenic route would be to overexpress the pathway producing the γ-glutamyl derivative of BIA, which is already present in *Lathyrus sativus* in order to divert metabolites away from the β-L-ODAP biosynthesis pathway (Kuo et al., 1998). There are at least two apparent problems with these approaches. Firstly, any system in which the enzymatic pathway of β-L-ODAP synthesis has not been disrupted in a way that makes the synthesis impossible is inherently unstable as any loss of expression of the transgene through mutation or transcriptional downregulation would result in the re-emergence of the toxin. Secondly, the legal regulatory issues associated with the introduction of a genetically modified crop, could severely delay the translation of β-L-ODAP-free grass pea genotypes into varieties, especially in the main grass-pea-growing countries, India, Bangladesh and Ethiopia. If β-L-ODAP-free genotypes of grass pea could be developed that do not fall under GMO-regulation, these could be made available to smallholder farmers much faster and with lower costs associated with varietal development and testing.

1.4.3 The importance of the genotype x environment interaction for β-L-ODAP production

Current evidence on the toxicity of β-L-ODAP and the socioeconomic context of grass pea consumption suggest that it is the toxin levels under stress conditions, not under normal conditions, that need to be considered to make grass pea safe. Four factors contribute to this effect. Firstly, stress conditions, in particular drought and flooding, are likely to affect all the crops in a regional farming system, causing crop losses and thus scarcity of food and money. This may result in an increased consumption of grass pea relative to other foods, as grass pea is more likely to withstand the stress conditions making it more available and affordable than many other foods (Girma et al., 2011; Butler et al., 1999). Secondly, the reduced food intake overall in poor households as well as the decreased diversity of food may cause malnutrition, leading to heightened susceptibility to lathyrism (Enneking, 2011). Thirdly, the β-L-ODAP levels in grass pea are increased under most stress conditions (Haque et al., 2011; Jiao et al., 2011a; Fikre et al., 2008; Jiang et al., 2013; Xiong et al., 2006), again resulting in increased toxin consumption. And fourthly, traditional methods of detoxifying grass pea for food may be compromised, in particular in times of drought, as the water and fuel necessary for prolonged steeping and boiling of grass pea seeds may be unavailable, leading people to switch to alternative and less safe methods of food preparation (Dufour, 2011; Girma et al., 2011).

These arguments are supported by the epidemiology of lathyrism. Outbreaks of the disease follow an epidemic fashion during or shortly after periods of extreme drought (Getahun et al., 1999; Haque et al., 1996; Simpson, 2002). Cases of lathyrism during normal environmental conditions are almost unheard of, even among the poorest (Girma et al., 2011). For these reasons, the toxin levels of released varieties under stress conditions are highly relevant for the safety of the crop, more so than the levels under normal conditions. Even high-toxin grass pea lines, such as most of the landraces currently cultivated by smallholders, are likely to be safe to consume, even as a significant part of the diet, as long as other food sources and detoxifying methods of food preparation are available (Girma et al., 2011). On the other hand, the factors described may compound to make even a normally low-toxin variety unsafe to consume in large amounts under particularly adverse conditions. Hence, the sensitivity of β-L-ODAP production to the environment and the mechanism by which it is regulated require closer attention and the generation of varieties that produce none or very little of the toxin, even under the most adverse conditions, should remain a major objective of breeding.

1.5   Towards the rapid domestication of grass pea

1.5.1   The problem of grass pea toxicity has not been fully resolved

Over the course of the past decade, several researchers have argued that the problem posed by β-L-ODAP in grass pea may have been overstated (Lambein, 2000; Rao, 2011; Lambein and Kuo, 2013). With the development of low-ODAP varieties, widespread knowledge of the risks of grass pea consumption and the emergence of more resilient food systems in some previously food insecure areas, new cases of neurolathyrism have become rare, leading to it being labelled "a disease of the past" (Singh and Rao, 2013).

Various natural toxins are present in commonly used food crops, such as solanine in crops of the nightshade family (e.g. potato, tomato and aubergine) (Willimott, 1933) or canavanine in several species of legumes (Bell et al., 1978), but their toxicity can be managed through correct storage and food processing procedures, so these food crops are not commonly thought of as poisonous. The perception of toxicity as a manageable risk may be more difficult to achieve for grass pea, as the crop has been strongly associated with famine and disease in the past.

A close parallel to the problem of neurotoxicity in grass pea is the toxicity of cassava (*Manihot esculenta*), caused by cyanogenic compounds (linamarin and lotaustralin) in the tubers. Despite the different chemical nature of these compounds, their consumption can result in a disease (konzo) with remarkable similarities to neurolathyrism – spastic paralysis of the legs that is often permanent – but can also result in acute cyanide poisoning and other health complications. Like grass pea, cassava is an important food security crop with high tolerance to drought. Cassava tubers, including genotypes with high cyanogen content (generally associated with bitterness) can be effectively detoxified using prolonged steeping and boiling (Cardoso et al., 2005). Low-cyanogen (sweet or cool) genotypes of cassava are available, but many smallholder farmers in South America and Africa prefer bitter varieties of cassava, as they are considered more durable in harsh climatic conditions and less susceptible to insect pests in the field and during storage of the harvested tubers (Chiwona-Karltun et al., 1998; Wilson and Dufour, 2002). The long preparation process necessary to detoxify the tubers is also considered to disincentivise theft. This preference has led to bitter cassava varieties persisting in agriculture despite the risk of toxicity and

the availability of very-low-toxin varieties. This highlights the importance of developing low-/zero-ODAP grass pea varieties that are suited to the needs of local farmers and testing them under realistic farming conditions.

The shift away from grass pea as a staple food and its inclusion in more diverse diets could help to build a new image of grass pea as a nutritious functional food, rather than a potentially toxic food of the poor. There are early signs that interest in grass pea may be on the rise again, especially for uses as green manure and as a fodder crop in China (Yang and Zhang, 2005), Australia (Siddique et al., 2006), the USA (Rao and Northup, 2011) and in Balkan countries (Mikic et al., 2011). In India there have been country-wide bans on the sale of food products derived from grass peas first implemented in 1961 and last reinforced in 2011 (FSSAI, 2011), as well as additional bans in several states. On 6 November 2015 the Food Safety and Standards Authority of India (FSSAI) issued a statement that the remaining bans on the sale of grass pea should be lifted, as the released low-toxin varieties can now be considered safe under most conditions, but this has proved controversial (Anand, 2016). In Ethiopia, where there has not been any outright ban on the cultivation or sale of grass pea, the crop has been increasing in acreage since 2000 (Haimanot et al., 2005; Girma and Korbu, 2012).

However, the spectre of toxicity associated with grass pea is still hampering efforts to improve this crop. Even if there is no evidence that reduced-ODAP varieties of grass pea still pose a risk of neurolathyrism, changing the thinking of research funding bodies, breeders, national agricultural authorities and individual farmers will be a very slow process as long as the possibility of toxicity persists. An important reason for this is that, to date, no animal model that accurately reflects the symptoms of human neurolathyrism as the result of oral consumption of β-L-ODAP has been developed (Mehta et al., 1983), as all existing animal models that show chronic hind leg paraparesis rely on injection of β-L-ODAP (Kusama-Eguchi et al., 2005; Kusama-Eguchi et al., 2010; Mehta et al., 1980; Mehta et al., 1976; Parker et al., 1979). In the absence of an appropriate animal model, it has been impossible to determine a safe level of β-L-ODAP consumption. Based on nutritional surveys, daily consumption of as much as 2 g of β-L-ODAP from grass pea may not cause any symptoms in humans not suffering malnourishment (Lambein and Kuo, 2013).

While there is no experimentally established threshold below which the concentration of β-L-ODAP in grass pea tissues can be regarded as safe, there are no known cases of human neurolathyrism from the consumption of only low-toxin grass peas. However the scarcity of

data on the usage of low-toxin grass pea varieties and on the epidemiology of neurolathyrism makes it difficult to draw any conclusions from this. This is further complicated by the apparent effect of nutritional status on the aetiology of lathyrism and the increased production of β-L-ODAP in the plant due to various abiotic stress factors. As argued above, for the full potential of grass pea as a food security crop to be realised, it needs to be made safe not only under normal conditions, when it forms part of a mixed diet, but also when populations are forced to rely heavily on grass pea in deficient diets.

Both the problem of toxicity under extreme circumstances and its perception as a dangerous food, which keeps grass pea from being improved and advocated in the first place, need to be resolved. The most promising approach to address both these aims remains to develop grass pea genotypes completely free of β-L-ODAP.

The year 2016 marks the UN Food and Agriculture Organisation's International Year of Pulses. In the words of the Global Pulse Confederation (an industry body), this year represents a "galvanizing moment to draw together key actors to further the contributions pulses make to health, nutrition, and sustainability" (GPC, 2016). Legume crops have great potential in meeting global food needs, especially due to their high protein content and low environmental impact, due to their symbiotic nitrogen fixation (Foyer et al., 2016). This project therefore makes a timely contribution to this goal. If the issue of toxicity in grass pea can resolved by developing genotypes with no β-L-ODAP or only safe levels of this toxin, regardless of the environmental conditions, it could kick-start the rapid domestication of grass pea from an orphan crop to a high-yielding, nutritious and sustainable crop for food and feed security in water-stressed areas.

1.5.2   From minor crop to major industrial product: the case of oilseed rape

An example of the usage of a minor crop species being revolutionised by the removal of its key limitations is the breeding of food-quality rapeseed (*Brassica napus*) varieties in the 1970s. Before this, the uses of rapeseed oil were limited to a few industrial applications, for example as a lubricant in marine engineering, as high levels of erucic acid, the most common fatty acid in rapeseed oil, and glucosinolates rendered the oil unpalatable. This picture changed dramatically with the introduction of low erucic acid varieties from 1970 and low glucosinolate varieties from 1975 (Kondra and Stefansson, 1970; Jönsson, 1977). These two changes allowed for the favourable nutritional characteristics of the oil and the

crop's high oil yield in temperate regions to be utilised for the production of food oils. In order to emphasise the different qualities of the new product and to set it apart from the limitations associated with rapeseed oil in the minds of many consumers, the new varieties were branded as "canola" (CANada Low erucic Acid). This re-branding was so successful that this name has virtually replaced the name of the crop in several countries. Between 1970 and 2000 the worldwide production of oilseed rape/canola increased from 5 Mt to 40 Mt (Baranyk and Fábry, 1999). Through the removal of the crucial factors limiting the acceptability of oilseed rape as a food crop, the advantages of this crop, including its high yield in temperate climates, and favourable fatty acid composition (Scarth and McVetty, 1999) were unlocked, rapidly making it the world's third largest oil crop in terms of production (FAO, 2013).

1.5.3   Approaches taken in this study

My objective in this PhD project was (i) to identify low- or zero-ODAP genotypes and mutants of grass pea and (ii) to better understand the synthesis of this compound in the crop, with the long-term goal of developing reliably safe varieties of grass pea. To this end, I used three approaches: the screening of existing grass pea germplasm for low-ODAP accessions; a mutant screen to identify low-/zero-ODAP plants from a mutagenised population of grass pea which might also facilitate the study of β-L-ODAP synthesis through forward genetics; a reverse genetic approach to identify genes involved in β-L-ODAP biosynthesis using sequence analysis and biochemical assays.

No entirely undomesticated *Lathyrus sativus* accessions are known, but considerable diversity exists among grass pea populations under cultivation. I therefore screened populations of grass pea germplasm with the aim of identifying accessions with reduced or zero-levels of β-L-ODAP. For this purpose, I adapted an existing spectrophotometric method to make it possible to screen large numbers of grass pea accessions for their variation in β-L-ODAP levels. I focused first on screening two populations ofgermplasm collected from around the world by seed banks in Germany and the United States. In collaboration with an Ethiopian researcher, I screened a larger population of landraces collected from Ethiopia and Eritrea, to identify diversity as yet untapped by breeders. The screening of grass pea germplasm for low-ODAP accessions is described in Chapter 2.

To increase the phenotypic diversity for β-L-ODAP content beyond the diversity present in germplasm collections, I screened an EMS-mutagenised population of grass pea. This method was appropriate for two reasons. Firstly, the ideotype (a grass pea mutant lacking β-L-ODAP) represents a loss of a naturally occurring function and should be possible to achieve by the inactivation of one or several genes. Secondly, very little is known about the genetics of toxin production or of grass pea in general, ruling out rational approaches to metabolic engineering. To be able to screen a large mutagenised population, I further streamlined the spectrophotometric method used in Chapter 2 to allow highly parallelised screening of a large mutant population. I used this to screen a population of over 3000 M2 families, a total of nearly 37000 plants, for low-/zero-ODAP mutants. From this population, I identified several mutant families with reduced β-L-ODAP concentrations in seedlings, shoots and seeds, but none that were entirely toxin-free. The method development for the mutant screen and the subsequent confirmation, characterisation and crossing of the identified mutants are described in Chapter 3.

To deliver on the goal of β-L-ODAP-free grass peas it may be necessary to first gain a better understanding of the genetic and biochemical basis of toxin production. To complement the forward genetics approach, I developed resources to enable a reverse genetics approach to investigate β-L-ODAP-biosynthesis. For this reason, I extracted and sequenced RNA from seven tissues at different developmental stages, with the help of The Genome Analysis Centre (TGAC). Based on the biochemistry of the proposed β-L-ODAP-biosynthetic pathway, and comparison to predicted protein sequences in related species, I identified a set of candidate genes for the enzyme catalysing the final step of synthesis. To test their functionality, I cloned the most promising of these genes, transiently expressed them in *Nicotiana benthamiana* leaves and supplied the tissue with the intermediates of the reaction. Transcriptome sequencing, sequence analysis and testing of candidate genes through heterologous expression are described in Chapter 4.

The implications of these discoveries for the development of improved varieties of grass pea and our understanding of β-L-ODAP synthesis are discussed in Chapter 5.

Chapter 2 – Screening grass pea germplasm for low-ODAP genotypes

2.1   Introduction

2.1.1   Diversity of grass pea germplasm

Grass pea is an ancient crop and both its benefits as a hardy food crop and the risk it may pose have been understood since antiquity (Erskine et al., 1994; Jackson and Yunus, 1984). Over 8000 years of cultivation, grass pea has been grown in many different areas around the world – the Balkans, South and Central Asia, East Africa, all around the Mediterranean, Central and Eastern Europe and, more recently, in China, Peru, Australia and Canada (Campbell, 1997). This long history of widespread cultivation implies that there might exist considerable genetic diversity in the primary gene pool of grass pea, which is supported by isozyme analysis (Gutiérrez-Marcos et al., 2006). Diversity can be observed for many phenotypes; flower colour, seed size, seed colour and patterning and biomass production (Tadesse and Bekele, 2003a; Vaz Patto and Rubiales, 2014; Polignano et al., 2005). This diversity is partially correlated with the geographic spread of grass pea cultivation. While most European accessions produce white flowers and large, flattened, cream coloured seeds, most South Asian accessions produce blue flowers and smaller, more rounded, dark and speckled seeds (Campbell, 1997). The most important issue holding back breeding efforts that would enable more widespread utilisation of grass pea as a food and fodder crop is the fact that it contains β-L-ODAP, a toxin believed to cause neurolathyrism in humans (Girma and Korbu, 2012; Yang and Zhang, 2005). If the diversity observed for other phenotypes mentioned above also exists for β-L-ODAP concentration in the seed, then low- or zero-toxin accessions may already occur within the existing germplasm of grass pea. Screening germplasm collections for such accessions therefore could provide the basis for the development of new, safer varieties.

Today, grass pea is mostly grown by smallholder farmers, as there is little formal trading in most countries (in part due to bans in India and China) and virtually no export market. Most farmers grow grass pea for their own consumption and as insurance for when their other crops fail due to its better resilience to unforeseeable weather extremes compared to

other legume crops, and especially its ability to withstand both drought and flooding (Campbell, 1997; Sarker et al., 2001). There have been some breeding efforts for grass pea in recent decades that have led to the release of several low-ODAP grass pea varieties in Bangladesh and India, but grass pea's status as a minor crop and the risk of human disease associated with it have limited the interest of breeders. Different regions have been claimed as the primary centres of diversity of grass pea, in particular South Asia (Wang et al., 2015; Chowdhury and Slinkard, 2000) and the Ethiopian highlands (Vavilov, 1927), but the putative centre of origin is in the Balkan peninsula based on archeobotanical evidence (Kislev, 1989), although grass pea is only very rarely grown in this region today (Mikic et al., 2011).

Grass pea germplasm from the Ethiopian highlands (a geographic area including much of Ethiopia, Eritrea and parts of Somaliland) has been studied primarily by Ethiopian scientists and breeders for over 50 years, but their attempts to improve grass pea genetically, in particular with regard to seed β-L-ODAP content have not had the success that was hoped for (Girma and Korbu, 2012). Yet it is likely that among the many local landraces of the area there exists untapped genetic diversity that could be useful for future breeding of grass pea. The Ethiopian highlands have been described as one of the five (later subdivided and expanded to twelve) principal centres of crop diversity by Vavilov (Vavilov, 1927). This region is an ancient agricultural centre and harbours a great amount of genetic diversity in cereal, vegetable and pulse crops, following thousands of years of selective breeding by smallholder farmers (Engels and Hawkes, 1991). The highly fragmented geography of the area, with many steep mountains more than 4000 m in height and narrow valleys with non-navigable rivers, may have impeded trade and cultural exchange, allowing considerable genetic diversity to develop among the crops cultivated across the region.

2.1.2   Breeding of low-ODAP varieties

Several breeding programmes have been undertaken to develop new low-ODAP varieties with seed ODAP concentrations below the (arbitrary) threshold of 0.1 % w/w). Most efforts have been based on germplasm from the Indian subcontinent. This has led to the release of the low-ODAP varieties BioL-212 (Ratan), Mahateora, LS 8246 and Prateek, as well as the medium-toxin varieties Pusa-24 and Nirmal (B1) (Campbell and Briggs, 1987; Dixit et al., 2016). The breeding relationships between these varieties are shown in Figure 5. The medium-toxin varieties Pusa-24 and Nirmal (B1) were derived by selection from Indian grass pea germplasm. The low-ODAP variety Ratan was derived from a selection of a callus

from a tissue culture population of Pusa-24 as part of a screen for somaclonal variation in ODAP content (Santha and Mehta, 2001). This variety was later crossed with the pink-flowered high-ODAP variety JRL-2 to create the pink-flowered low-ODAP variety Mahateora (Sastri, 2008). The variety LS8246 was derived by selection from Pusa-24 and its low-ODAP phenotype is likely be due to a naturally occurring mutation. A population of 64 single-seed selections from Pusa-24 were grown at Morden, Manitoba, Canada. These plants were selfed and seeds produced from offspring plants were assayed for ODAP. One plant was selected as having a very low ODAP content. Seed from this plant was bulked to give rise to LS8246 (Campbell and Briggs, 1987). LS8246 was crossed with the high-ODAP Indian landrace a-60 tolerant to powdery mildew to give rise to the variety Prateek (Sastri, 2008). Low-toxin varieties have also been released in Bangladesh under the names BARI Khesari 1, 2 and 3 (Akter et al., 2015). One low-toxin variety, Ceora, has been developed in Australia (Siddique et al., 2006). Only one improved variety of grass pea (Wasie) has been released in Ethiopia to date (Kumar et al., 2011a).
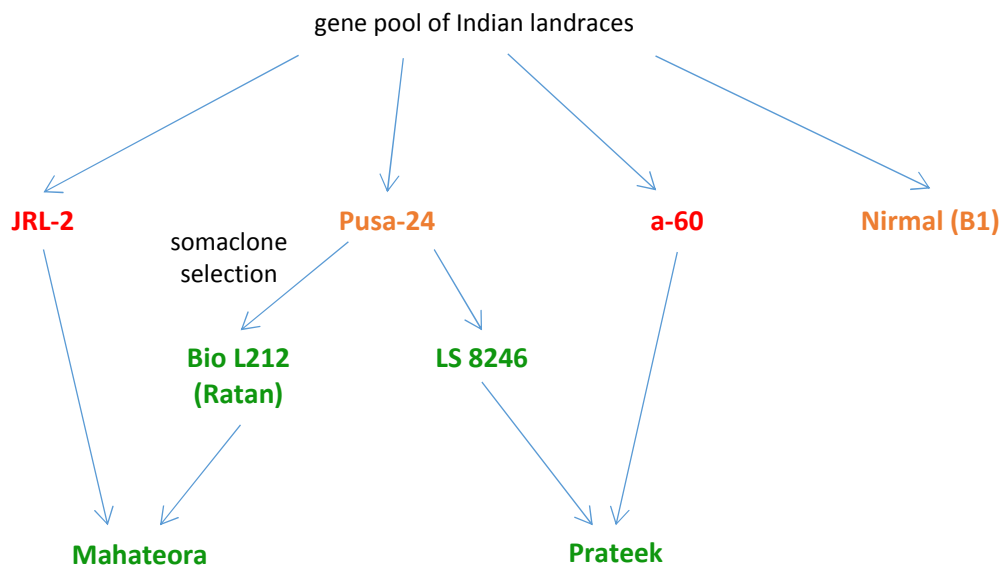
Figure 5. Relationships between Indian grass pea varieties (shown in bold). Low-toxin varieties are shown in green, medium-toxin varieties are shown in orange, high-toxin varieties are shown in red.

These varieties represent an important advance in the improvement of grass pea. Their reduced toxin levels reduce the risk of neurolathyrism among consumers. However, none of these varieties have achieved the goal of an entirely β-L-ODAP-free crop. The ODAP content of existing "low-toxin" varieties can increase to potentially dangerous levels if the

plants are grown under adverse environmental conditions (Fikre et al., 2008). However, it is conceivable that among the existing genetic diversity there are accessions with low seed ODAP contents that have not been assayed previously. Such accessions could form the basis for future breeding programmes to develop novel low-toxin varieties. Depending on the genetic basis for the low toxin phenotypes, it is possible that alleles from yet unknown low-toxin accessions could be crossed into the background of existing low-toxin genotypes to reduce further β-L-ODAP levels in the plant or to eliminate the toxin entirely. I therefore assayed collections of grass pea germplasm to identify previously unknown low-toxin accessions. To access as much diversity as possible, I focused on international collections of germplasm comprising accessions collected from grass pea cultivating countries around the world, as well as a population of landraces collected from the Ethiopian highlands region.

2.1.3   Variation in seed β-L-ODAP concentrations between individual seeds and between growth environments

The conditions of cultivation appear to have a significant impact on the amount of β-L-ODAP a grass pea plant produces (Xiong et al., 2006; Fikre et al., 2011; Jiao et al., 2011a; Xing et al., 2001). This is important agronomically, as the amount of ODAP present in grass pea seeds may become dangerously high under some conditions during the growing season, in particular late-season drought (Fikre et al., 2011). This sensitivity of ODAP production to growth conditions may complicate experiments to compare grass pea genotypes, if material is sampled from plants grown in different environments. As I have argued in section 1.4.3, the β-L-ODAP-levels present in a grass pea plant under conditions of environmental stress may be more relevant for the selection of safe genotypes than β-L-ODAP-levels under non-stressed conditions. To estimate the extent of variation in β-L-ODAP-levels in the seeds of grass pea plants grown under different experimental conditions, I tested seed batches of five Indian grass pea varieties (LSWT11, Mahateora, Ratan, Pusa-24 and Nirmal) grown in field conditions in West Bengal, India and in glasshouse and field conditions at JIC. For this experiment, I measured the β-L-ODAP-concentrations in five individual seeds from each batch, allowing me to measure the variation between individual seeds. I used these data to inform the design of the experiments to measure β-L-ODAP concentrations in the grass pea germplasm collections.

2.1.4   Methods for measuring ODAP concentrations

Several methods have been described to measure the amount of β-L-ODAP in grass pea tissues. The Rao method, a spectrophotometric assay (Rao, 1978), named after its inventor

S.L.N. Rao, uses o-phthalaldehyde (OPA) and β-mercaptoethanol and was the first reliable quantitative method to measure ODAP-levels in grass pea tissues. Its accuracy and precision have been improved through later modifications by other researchers (Briggs et al., 1983; Hussain et al., 1994). More recent methods to measure ODAP concentrations in grass pea make use of high performance liquid chromatography (HPLC) (Yan et al., 2005) or capillary zone electrophoresis (CZE) (Arentoft and Greirson, 1995). Techniques using liquid chromatography mass spectrometry (LCMS) (Koh et al., 2005) or gas chromatography mass spectrometry (GCMS) (Xie et al., 2007) have been described to measure ODAP in ginseng (*Panax spp.*), but have not yet been applied to grass pea. These methods offer better accuracy and sensitivity than the Rao method, but they are more costly and time consuming, as they require samples to be processed one-by-one. The spectrophotometric method has been the most widely used assay for ODAP in grass pea samples, because it relies on simple equipment that can be found in most laboratories, even those with few resources (Tadesse and Bekele, 2003b; Campbell, 1997; Srivastava and Srivastava, 2006).

Several variations of this method have been described in the literature (Rao, 1978; Briggs et al., 1983; Hussain et al., 1994), but all are based on the same basic reaction of 2,3-diaminopropionic acid with β-mercaptoethanol and o-phthalaldehyde in the presence of a tetraborate buffer system to form a yellow soluble pigment. The concentration of the pigment can be estimated by measuring absorbance at 420nm using a spectrophotometer. Hussain et al. have published a reaction scheme for this reaction (redrawn in Figure 6A) (Hussain et al., 1994). The reaction products shown in this publication appear unlikely, as in each of the two products one of the carbon atoms in the six membered rings with delocalised electrons is also shown to have four covalent bonds. In addition, the reaction as shown would imply the loss of seven hydrogen and two oxygen atoms per reaction between the substrates and the products, although no production of gas is observed in the reaction. Figure 6B shows my proposed correction of the structural formulae of the products of this reaction. The corrected reaction scheme implies the loss of four hydrogen and two oxygen atoms (i.e. two water molecules) per reaction.
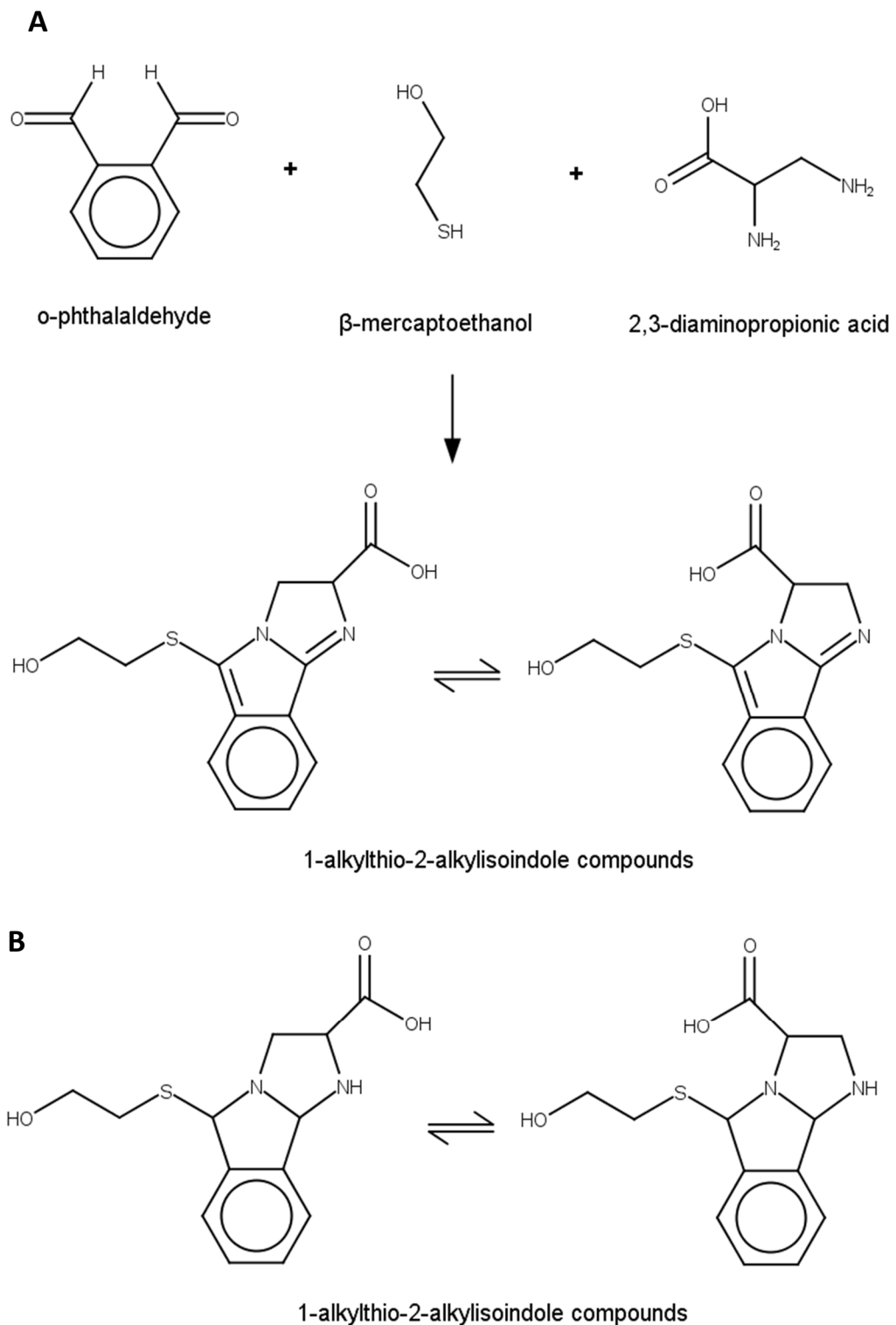
Figure 6. Colour forming reaction used in the spectrophotometric assay. The 1-alkylthio-2alkylisoindole compounds have a characteristic absorption peak at 420 nm. A) reaction scheme as given by Hussain et al. (1994) B) Proposed corrected reaction products

The 1-alkylthio-2-alkylisoindoles that are produced by this reaction have a very different absorption spectrum to the monoamino acids that also react with OPA/β-mercaptoethanol (Simons Jr and Johnson, 1976), allowing for low background readings. While more accurate methods to measure ODAP concentrations are now available, this method still provides the most scalable protocol. The chemistry of this assay provides flexibility when processing large numbers of samples as all reactions are end-point reactions and all chemicals and materials are stable at room temperature for the duration of the assay.

Building on the adjustments to the spectrophotometric method made by Briggs et al. (Briggs et al., 1983) I scaled down the chemistry used for the final, colour-forming reaction to allow measurements in plate format. This method was used to assay the ODAP-concentrations in the germplasm collections. I assayed the Ethiopian grass pea germplasm collection together with Alemu Abate, an Ethiopian researcher from Aksum University, Ethiopia, during a placement at the Biosciences Eastern and Central Africa (BecA) Hub in Nairobi, Kenya, because of the difficulty of accessing this germplasm. The low cost and equipment requirements of the spectrophotometric assay will allow my collaborator to apply the spectrophotometric ODAP assay to germplasm screening and pre-breeding projects at his home institution.

2.2   Materials and methods

2.2.1   Germplasm

Three different sources of germplasm were used to assess the natural diversity of ODAP levels present in the global grass pea population. The first was a population collected by the Institut für Pflanzenkunde (IPK) in Gatersleben, Germany, which contained 44 lines of grass pea collected from several countries, mostly from south and south-east Europe. To add to this population, 96 grass pea accessions were requested from the United States Department of Agriculture at Pullman, Washington, USA, which maintains an international collection of many crop species. To increase the overall diversity of the population, accessions from countries that were not represented in the IPK population were selected from the population held at the USDA seed bank. Where possible, local landraces, rather than varieties, were selected in order to capture local variation. However, the passport information associated with some of the accessions was incomplete, meaning that the exact origin was not known for all. Detailed lists of accessions obtained from each population along with their geographic origins are provided in Appendix 1.1.3.

Further attempts were made to complement these grass pea germplasm collections by contacting the seed collections held by the International Centre for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria and the Université de Pau in France, the two largest collections of grass pea accessions. However, no seeds were forthcoming from either source. The ICARDA seed bank has been unable to distribute seeds as their main seed store at Aleppo has been disrupted by the Syrian civil war, although their seed collection has been duplicated at other centres. The Université de Pau collection has ceased active distribution and is now held at the Conservatoire Botanique National des Pyrénées et de Midi Pyrénées (CBNPMP), where it is not being regularly regenerated (personal communication, Jocelyne Cambecèdes, CBNPMP, France), placing available grass pea diversity at risk.

The accessions received from the two germplasm collections, along with the varieties provided by BCKV, India, covered virtually the entire range of countries in which grass pea is grown (see Figure 7).
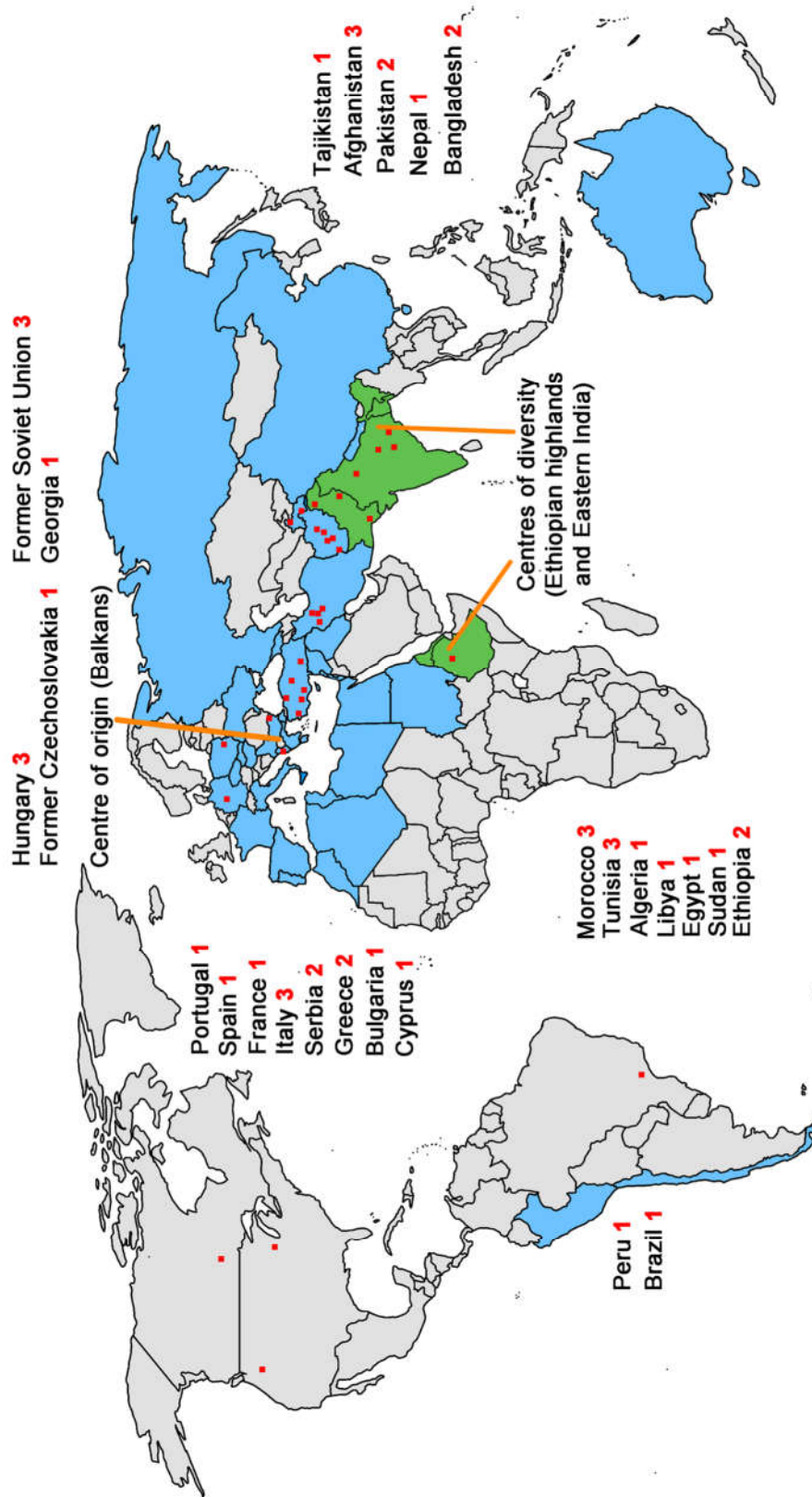
Figure 7. Map of origins of the grass pea accessions obtained from the IPK seed bank, Gatersleben, Germany and the USDA seed bank, Pullman, Washington, USA. Accessions with known coordinates of origin are represented as red squares. In addition, the numbers of accessions for which only the country of origin, but no coordinates of the collection site are recorded in the accession metadata are shown in red figures next to the name of their country of origin. Major grass pea cultivating countries are shaded in green, minor cultivating countries are shaded in blue based on Campbell, (1997), other countries are shaded in grey. Centres of diversity and the putative centre of origin for grass pea are marked.

A collection of 413 Ethiopian grass pea landraces was previously assembled by the Debre Zeit Research Centre (DZRC) under the supervision of the Ethiopian Institute for Agricultural Research (EIAR). Seeds were collected from a total of 85 districts (woredas) in the regions Oromia (western, central and southern Ethiopia), Tigray (northern Ethiopia) and Amhara (northwestern Ethiopia), as well as one district from the Somali region of Ethiopia. Three additional accessions in this population were collected from southern Eritrea before the country's independence from Ethiopia in 1993. This population was supplemented by 10 grass pea accessions from Australia, which had been donated to the Ethiopian Biodiversity Institute (EBI).

2.2.2   Production of seed material for germplasm assays

Seeds of the accessions obtained from IPK and USDA collections and of the Indian grass pea varieties LSWT11, Mahateora, Nirmal, Ratan and Pusa-24 were sown in rows in the field at JIC (five seeds per accession). The plants were grown up a wire-mesh support for climbing inside a protective cage to exclude birds, rabbits and rodents. The plots were irrigated as required by rainfall patterns, but not fertilised. Plants of the Indian varieties were also grown in glasshouses at JIC during February to June and during September to January. The shoots of mature plants were harvested, air-dried and threshed to release the seeds. The seeds of each accession were pooled at the threshing stage. For cultivation in the glasshouse, seeds were scarified by rubbing with sandpaper and imbibed in water for 48 h before being sown into 10-litre rose pots, with six plants per pot. During winter and autumn months plants were grown under mercury halide (Hydrargyrum Quartz Iodide, HQI) lamps to ensure at least 12 h of light every day. The glasshouses were also heated to prevent temperatures from falling below 15 °C. Plants were left to mature and dry in the field or in their pots before being harvested and threshed. After this, seeds were stored at 4 °C in dry storage before being used in the assay. The EIAR population was maintained at DZRC and all the seeds used for ODAP-testing were grown in field plots at the DZRC, Bishoftu (Ethiopia). After harvest, the seeds were stored in a dry seed-storage unit at 4 °C.

2.2.3   Comparison of extraction media

For water extraction, ground seed meal (20 mg) was weighed out and suspended in 5 ml of deionised water, followed by incubation in a water bath at 95 °C for 15 minutes. For ethanol extraction, ground seed meal (20 mg) was suspended in 5 ml 60 % v/v ethanol in deionised water, followed by overnight incubation in an orbital shaker at room temperature. In both cases, five replicate extractions were performed. After extraction,

samples were centrifuged at 3100 g for 30 min in an Eppendorf 5810R centrifuge (Eppendorf, Hamburg, Germany). The supernatant (100 µl) was mixed with 200 µl of 3M KOH solution and incubated at 95 °C for 30 min. The reaction product (300 µl) was transferred to a spectrophotometer cuvette, diluted with 700 µl of deionised water and mixed with 2 ml of reaction buffer (7.5 mM o-phthaldialdehyde (OPA), 0.5 M potassium tetraborate tetrahydrate, 1 % v/v ethanol, 0.2 % v/v β-mercaptoethanol) by tapping the cuvette on the bench. Reactions were incubated at room temperature for 30 min. Absorbance was measured using a cuvette spectrophotometer (Lambda Bio, PerkinElmer, Waltham, Massachusetts, USA)

2.2.4   Sensitivity and linearity testing of the plate-based spectrophotometric assay

Eight samples of ground seed meal of variety LS007 (sample weights ranging from 10.7 mg to 17.7 mg) were suspended in 1 ml of deionised water and extracted by incubating in a water bath at 95 °C for 30 minutes. Separately, a 3.2 mM solution of β-L-ODAP standard (Lathyrus Technologies, Hyderabad, India) in 3M KOH was prepared. This solution was serially diluted twofold with 3M KOH ten times to produce a dilution series from 0.0031 mM to 3.2 mM. These solutions were incubated at 95 °C for 30 minutes to hydrolyse β-L-ODAP to L-DAP. In a 96-well, flat-bottom plate (Greiner Bio-One, Alphen aan Den Rijn, Netherlands) suitable for use in plate spectrophotometers, one aliquot of 10 µl of each of the supernatants was mixed with 20 µl of each of the standard dilutions in 3M KOH or 3M KOH with no L-DAP. A separate plate was prepared containing eight replicates of 20 µl of the L-DAP standard solutions mixed with 10 µl of deionised water.

O-phthalaldehyde/tetraborate reaction buffer was prepared by dissolving 5.29 g of potassium tetraborate tetrahydrate (Sigma-Aldrich, St. Louis, Missouri, USA) in 42 ml of dH$_2$O, by shaking and gently warming the mixture. Separately, 34.6 mg of o-phthalaldehyde (Sigma-Aldrich) were dissolved in 346 µl of absolute ethanol and 69 µl of β-mercaptoethanol (Sigma-Aldrich). Once both o-phthalaldehyde and potassium tetraborate were fully dissolved, both solutions were combined. The resulting 42.4 ml of reagent buffer were sufficient for processing 96 samples in microtitre plates, including both hydrolysed and non-hydrolysed aliquots. This buffer is more dilute than the reagent buffer used by Briggs et al. (1983), who diluted sample extracts with water prior to the addition of the reagent buffer. The final concentration of reagents in the reaction was the same in the experiments described here and in the method developed by Briggs et al. (Briggs et al., 1983; Hussain et al., 1994).

For the colour-forming reaction, 220 µl of OPA/tetraborate buffer were added to each well of the plates containing samples and standards. Plates were left to incubate at room temperature for 25 minutes. Absorbance at 420 nm was measured using a plate spectrophotometer (VersaMax, Molecular Devices, Wokingham, UK).

2.2.5   Spectrophotometric assay for ODAP concentrations in individual seeds

To measure the ODAP concentrations of seed batches produced in different settings, individual seeds were assayed. Seeds were clamped using a rubber-padded peg and drilled using a benchtop drill (Xenox, Föhren, Germany) to extract seed meal from the storage cotyledons as shown in Figure 8. Care was taken to avoid the embryonic axis of the seed by drilling into the flattened side of the seed, which is opposite the embryonic axis. This was to allow the seed to germinate later to produce progeny, if needed.

Figure 8. Drilling to collect seed meal material from the storage cotyledons of individual seeds. Care was taken to avoid drilling into the embryonic axis of the seed, allowing the seed to be sown after drilling, if needed. The seed was held in place during drilling using a peg.

Seed meal was collected in a microcentrifuge tube and dried for 48 hours in a freeze-drier (BenchTop SLC®, Virtis, Gardiner, New York, USA). Dried meal samples were weighed and

the weights recorded electronically to allow later normalisation. A scaled-down version of the spectrophotometric method as modified by Briggs et al. (1983) was used to measure ODAP concentrations in seed samples. Volumes and sample weights were scaled down for extractions and reactions to take place in plate format. This allowed parallelised processing of samples using multi-channel pipettes and rapid measurement using a plate-spectrophotometer.

Free amino acids (including L-DAP and ODAP) were extracted from the seed meal; 600 µl of 60 % v/v ethanol in distilled water were added to each sample, followed by incubation at room temperature in a shaking incubator for 22 h. Samples were centrifuged at 16,250 g in a benchtop centrifuge (Biofuge Pico, Heraeus, Hanau, Germany) for 10 minutes. An aliquot (80 µl) of the supernatant of each extracted sample was transferred into a 96-well microtitre plate (Sterilin, Newport, UK), and 160 µl of 3M KOH solution were added to each well. Plates were sealed firmly and clamped tightly between aluminium plates to prevent leakage. The plates were submerged in a water bath at 95 °C for 30 minutes to hydrolyse ODAP to L-DAP. After this, the plates were submerged in water at room temperature to cool them down before drying and releasing the clamps, to prevent the ethanol in the solution from boiling off.

OPA/tetraborate buffer was prepared as described in section 2.2.4. For the colour-forming reaction, 30 µl of hydrolysate was mixed with 220 µl of OPA/tetraborate buffer in a 96-well microtitre plate with a clear flat bottom (Greiner Bio-One, Alphen aan Den Rijn, Netherlands). Separately, another flat-bottom plate was loaded with 20 µl of 3 M potassium hydroxide solution and 10 µl of non-hydrolysed supernatant from the overnight extraction, immediately followed by 220 µl of OPA/tetraborate buffer.

The reaction mixture in each well was mixed by gentle sideways tapping of the plate, followed by incubation at room temperature for 30 minutes. Absorbance at 420 nm was measured using an optical plate reader (Indian varieties, IPK population and USDA population – VersaMax, Molecular Devices, Wokingham, UK; EIAR population – FLUOStar® Omega, BMG Labtech, Ortenberg, Germany).

2.2.6   Spectrophotometric assay for ODAP in bulk seed samples

To measure seed-ODAP concentrations of grass pea accessions directly from germplasm collections, bulk meal samples of milled seeds were used. Dry seed samples (5 g of IPK and USDA population seeds; 9 g of EIAR population seeds) were weighed out and ground for ca.

1 minute using a coffee mill (IPK population – BISTRO electric coffee grinder, Bodum, Triengen, Switzerland; USDA population – F20342 Coffee mill, Krups, Solingen, Germany; EIAR population – CBM4 Coffee and Spice Grinder, Black+Decker, Towson, Maryland, USA) and collected in a 50 ml Falcon tube. Between millings, the coffee mills were wiped clean using 95 % v/v ethanol and allowed to dry. Dry ice was inserted into the mill between millings every 3-5 samples to stop the mill from overheating. Dry ice was removed before adding the next sample.

For the germplasm screen, the extraction of ODAP from ground seed samples and hydrolysis was performed as described by Briggs et al. (1983), but the colour-forming reaction was performed in plate-format, using scaled-down chemistry to allow automated measurement using an absorbance plate-reader. An amount between 0.4 g and 0.5 g of seed meal from air-dried seeds was weighed out for each sample and the exact weight recorded. To each sample, 10 ml of 60 % v/v ethanol in dH_2O were added to extract free amino acids. Samples were left to extract with the tubes positioned horizontally on an orbital shaker at room temperature for between 22 h and 36 h. After this, samples were stored upright at 4 °C for up to 24 h until used in the assay. Samples were centrifuged for 15 minutes at 3100 g in a Eppendorf 5810R centrifuge (Eppendorf, Hamburg, Germany) to separate the extracted seed meal from the supernatant containing the extracted amino acids. Of this supernatant, 2 ml were mixed with 4 ml of 3M KOH solution in dH_2O. This reaction mixture was incubated in tightly capped 15 ml Falcon tubes for 30 minutes at 95 °C to hydrolyse ODAP to L-DAP. Tubes were chilled back to room temperature before opening to prevent the boiling off of ethanol from the solution. Following hydrolysis, the colour forming reaction using the OPA/tetraborate reaction buffer was performed in 96-well flat-bottomed microtitre plates as described in section 2.2.4. Three technical replicates were prepared for each hydrolysed sample, along with one plate containing 10 µl of non-hydrolysed extract, 20 µl of 3M KOH and 220 µl of OPA/tetraborate buffer.

2.2.7   Calculation of ODAP concentrations from absorbance readings

The 420 nm absorbance readings of non-hydrolysed samples were subtracted from the readings of the hydrolysed samples to exclude absorbance caused by the polystyrene plate bottom, the extraction buffer, potassium hydroxide solution and reagent buffer as well as other compounds extracted from the seed tissue. The difference between the two values

served as a measure for ODAP concentration. The dry weight of samples, which had been recorded before the extraction, was used to normalise the results.

A set of eleven (individual seed samples, IPK population bulk samples and USDA population bulk samples) or seven (EIAR population bulk samples) linear dilutions of L-DAP.HCl (Sigma-Aldrich, St. Louis, Missouri, USA) was included on each measurement plate as a positive control of the colour forming reaction and to establish a standard curve for calibration of the assay. Aliquots (30 µl) of known concentrations of L-DAP.HCl in 2 M KOH solution were dispensed into the flat-bottom plates. OPA/tetraborate reagent buffer (220 µl) was added to each L-DAP.HCl standard and to one well containing 30 µl of 2M KOH solution with no L-DAP.HCl. Concentrations of L-DAP.HCl were chosen to cover the expected range of equivalent seed ODAP concentrations. Standard concentrations used alongside the experiments testing single seeds from different seed batches, and bulk samples from the three germplasm collections, are shown in Appendix 1.1.

Linear regressions of standard curves were calculated using the LINEST function in Microsoft Excel for Windows 2013. The same software was used to calculate seed ODAP concentrations using the following formula:

$$conc = \frac{\left(A_{hyd} \quad A_{non-hyd}\right) \times V_{ext}}{m_{sample} \times a_{standard}} \times 100\,\%$$

Conc	seed ODAP concentration in % w/w
A$_{hyd}$	absorbance reading of hydrolysed sample after the colour-forming reaction (averaged across technical replicates)
A$_{non-hyd}$	*absorbance reading of non-hydrolysed sample*
V$_{ext}$	*Volume of extraction buffer*
m$_{sample}$	*Mass of the seed meal sample*
a$_{standard}$	*Slope of the standard curve*

The software was also used to calculate means and standard errors and to plot bar charts and scatterplots representing the data. Histograms were plotted using the programming environment RStudio (version 0.98.1060, https://www.rstudio.com/), running the statistical programming language R (version 3.1.2, https://cran.r-project.org/bin/windows/).

2.3   Results and discussion

2.3.1   Optimisation of the spectrophotometric method

Two media have been described for extracting ODAP from grass pea tissues for the spectrophotometric method: water and 60 % v/v ethanol in water. For extraction in water, samples are incubated at 95 °C for 15-30 minutes while 60 % v/v ethanol extractions are done at room temperature overnight. To decide which medium to use for my experiments I performed the spectrophotometric assay using five extractions each, using water or 60 % v/v ethanol on samples of LS007 seed meal. The spectrophotometric assay was performed as described in section 2.2.4. Water extraction produced absorbance readings with mean 0.161 ± 0.011 (standard error). Extraction with 60 % v/v ethanol produced absorbance readings of mean 0.154 ± 0.007 (standard error). These results are also shown in Figure 9.
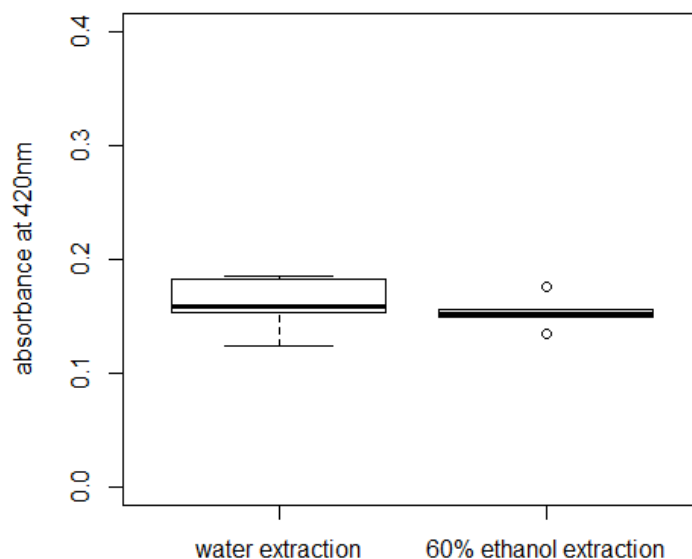
Figure 9. Absorbance measurements at 420nm of the spectrophotometric assay based on grass pea seed meal samples (20 mg) extracted using water or 60 % v/v ethanol in water as extraction media. Boxes show lower and upper quartiles surrounding the median, whiskers show the most extreme datapoint within 1.5 times the interquartile range from the box. Circles show outliers beyond that range.

Both extraction methods produced similar mean results, but the extraction method using 60 % v/v ethanol resulted in slightly lower variation between the replicates. A problem with using water as the extraction medium became apparent while processing these samples. After the incubation at 95 °C the sample suspension had taken on a thick, gel-like consistency, which prevented the seed meal and the supernatant from properly separating

during centrifugation. This problem did not occur when using 60% v/v ethanol, nor when using water for extraction of ODAP from samples other than seed meal. A different problem occurred when using 60 % v/v ethanol for extraction of ODAP from green tissues, as ethanol also extracted chlorophyll, which could lead to unwanted background noise in the spectrophotometric assay. For this reason, I decided to rely on 60 % v/v ethanol for the extraction of ODAP from seed meal samples, but use water for the extraction of ODAP from other tissues, as described in the next chapter.

To test the linearity and sensitivity of the plate-based spectrophotometric assay I compared the absorbance caused by non-hydrolysed seed meal extract spiked with serial dilutions of L-DAP and the L-DAP standard series alone. I extracted free amino acids from eight samples of grass pea seed meal (variety LS007, sample weights ranging from 10.7 mg to 17.7 mg). These samples did not undergo hydrolysis treatment. Aliquots of the extract were spiked with serial dilutions of β-L-ODAP (Lathyrus Technologies, Hyderabad, India) in 3M KOH, which had been hydrolysed to produce L-DAP. This allowed me to measure the background absorbance caused by compounds other than β-L-ODAP in the extract. As shown in Figure 10, the serial dilutions of L-DAP produced absorbance values that increase linearly with L-DAP concentration in the range of 0.025 mM to 3.2 mM (blue curve). These molar concentrations of L-DAP were equivalent to extracts of grass pea samples of 15 mg with β-L-ODAP concentrations ranging from 0.018 % to 2.25 % w/w of dry weight. At L-DAP concentrations lower than 0.025 mM, the absorbance measurements showed great variation between readings, indicating the detection limit of this assay. The non-hydrolysed grass pea extracts with added L-DAP (which is equivalent to hydrolysed extracts from grass pea samples with varying β-L-ODAP concentrations) produced significantly higher readings than the L-DAP standards at low L-DAP concentrations (green curve). This shows that blanking was necessary to remove the background absorbance caused by other compounds in the extract. After subtracting the absorbance measurements of the non-hydrolysed extracts alone, the readings of the blanked spiked extracts (orange curve) became indistinguishable from the readings of the dilution series for L-DAP alone. This showed that the plate-based assay is sufficiently sensitive and accurate to allow the measurement of β-L-ODAP in grass pea samples containing β-L-ODAP in the range indicated by the literature (Tay et al., 2000; Sharma et al., 2000; Campbell, 1997; Sarwar et al., 1995; Deshpande and Campbell, 1992). The measurement of absorbance produced by non-hydrolysed samples was necessary to correctly blank readings and produce accurate measurements, especially at low concentrations of L-DAP/β-L-ODAP.
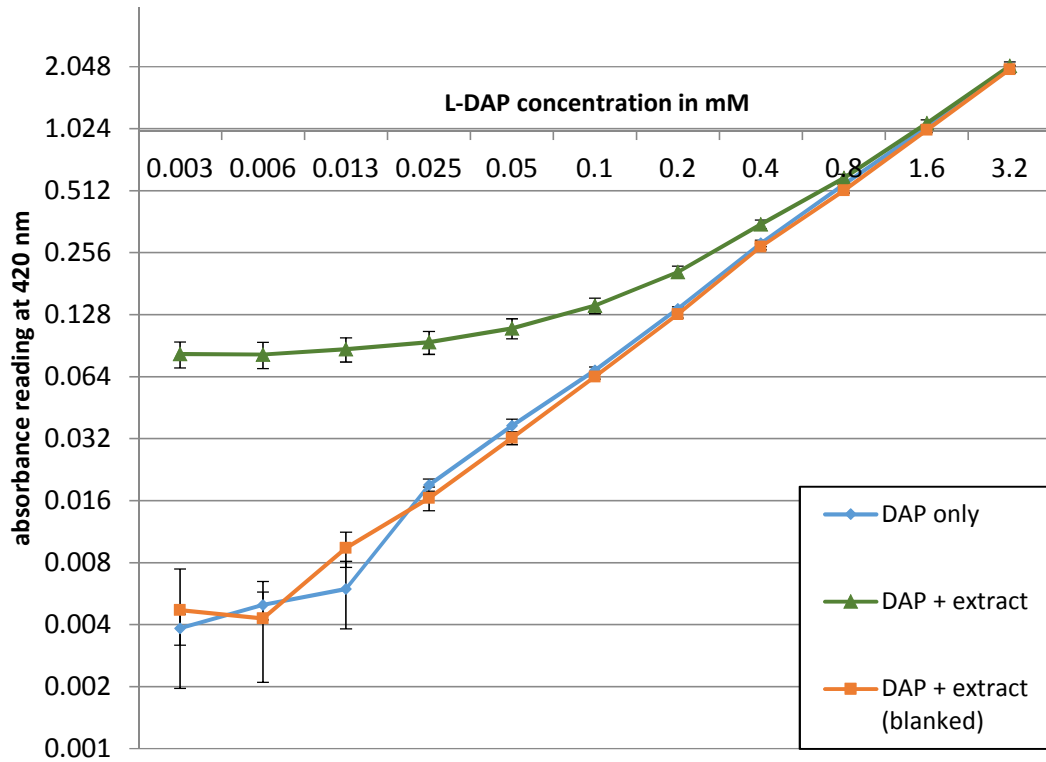
Figure 10. Linearity of absorbance values produced by the spectrophotometric method. blue: serial dilution of DAP; green: serial dilution of L-DAP and non-hydrolysed extract without blanking; orange: serial dilution of DAP and non-hydrolysed extract, blanked to remove the background absorbance present in non-hydrolysed extract without L-DAP. Each datapoint represents an average of eight samples, error bars denote standard error. Both axes are logarithmic (base 2).

ODAP occurs in two isomers, an α- and a β-form, depending on to which of the two amino groups the oxalyl-residue is attached. One limitation of the spectrophotometric method of measuring ODAP is its inability to differentiate between the α- and β- isomers of ODAP (Hussain et al., 1994). This is because of the hydrolysis step of the assay in which both α- and β-L-ODAP are converted into the same compound, L-2,3-diaminopropionic acid (L-DAP). Only the β-isomer appears to cause lathyrism, while the α-isomer is non-toxic (Chase et al., 1985; Wu et al., 1976). Both isomers are produced when the molecule is synthesised chemically, but the biosynthesis of the compound in the plant appears to differentiate strongly between the two forms, leading to 95% of ODAP present being the toxic β-isomer (Roy and Rao, 1968). The two isomers spontaneously interconvert via a reversible, non-enzymatic reaction (see Figure 11), leading to an equilibrium of 60% β-isomer and 40% α-isomer at room temperature (Abegaz et al., 1993). This reaction occurs slowly at room temperature and at neutral pH, but faster at alkaline or acidic pH and higher temperatures (Padmajaprasad et al., 1997).
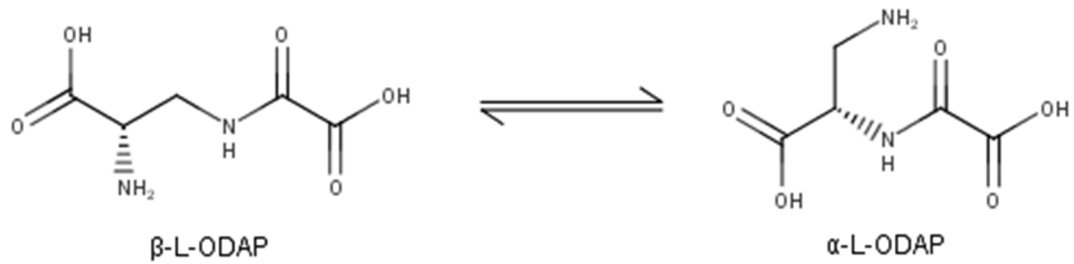
Figure 11. Spontaneous isomerisation between α- and β-ODAP

The production of α-L-ODAP instead of β-L-ODAP has been proposed as a breeding target for grass pea (Campbell, 1997), because it was hoped that this might retain any positive effects the compound might have on the physiology of the plant while removing its toxic effects on mammals. However, no germplasm accumulating an increased proportion of α-L-ODAP has been identified. In addition, it must be remembered that grass pea seeds and shoots are usually prepared for human consumption by steeping and boiling in water (Enneking, 2011; Dufour, 2011), methods that are traditionally used to destroy ODAP. Normally this leads to the partial detoxification of the tissue as some of the β-L-ODAP is converted into the non-toxic α-isomer, approaching a pH- and termperature-dependent equilibrium (Padmajaprasad et al., 1997). This processing method would have the opposite effect on high-α-/low-β-L-ODAP tissues on the other side of this equilibrium, leading to the production of β-L-ODAP. For this reason, I decided not to try to differentiate between the two isomers while conducting the mutant screen, but rather to use the spectrophotometric method, which effectively measures total ODAP concentration in a sample, despite this limitation.

2.3.2   Seed ODAP variation between batches grown in different conditions

To test whether environmental effects are likely to impact the toxin levels measured in the seeds of the grass pea accessions used in this study, I measured the toxin concentration in individual seeds of 6 genotypes: LSWT11, LS007, Pusa-24, Nirmal, Ratan and Mahateora grown in different environments.

I measured the ODAP concentrations in five individual seeds of different seed batches grown in a glasshouse and in the field at the JIC as well as seeds directly provided by BCKV and, in the case of one genotype (Pusa-24), by the USDA seed bank, using the methods described in sections 2.2.4 and 2.2.7. The results of this experiment are shown in Figure 12. These data showed remarkable differences between batches of the same accessions grown in different settings. In particular, seeds grown in a heated glasshouse at the JIC during

winter 2015/2016 (batch 'JIC glasshouse 2016') contained lower concentrations of ODAP than seeds grown at the JIC in a glasshouse during summer 2015 (batch 'JIC glasshouse 2015'). This difference was significant (p < 0.01) for the varieties Pusa-24 and Mahateora. While this did not provide a measurement of the impact of any one specific environmental factor on ODAP-production, it gave an indication of the variability that could be expected from different experimental conditions.

Only seeds of the variety Mahateora contained consistently less than 0.1 % w/w ODAP, the commonly applied, albeit arbitrary, threshold (Asthana, 1995; Chakrabarti et al., 1999) for low-toxin varieties. Despite being classed as a low-toxin variety, seed batches of Ratan (BioL-212) displayed an averageODAP concentration much higher than this threshold.
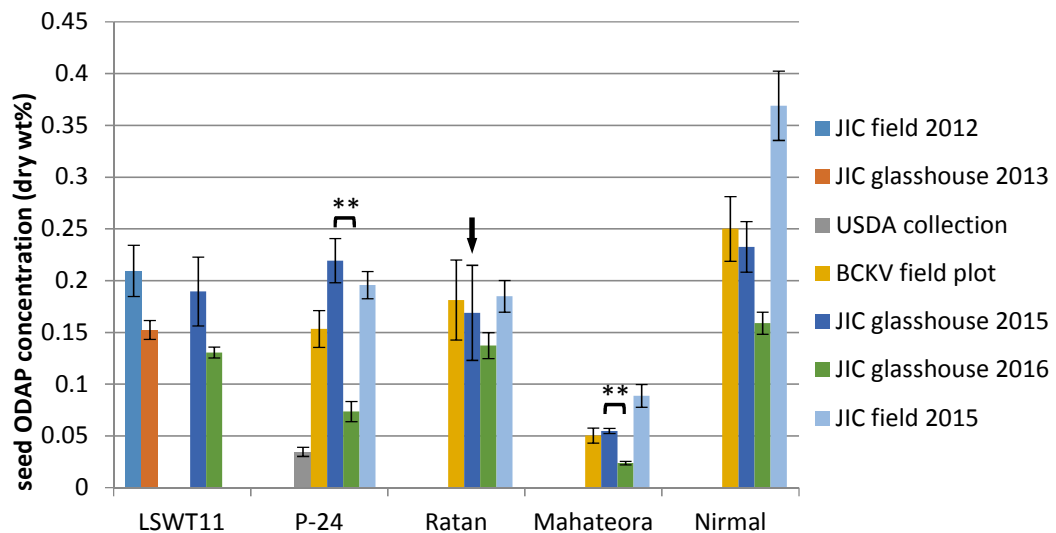
*Figure 12. Seed ODAP concentrations in grass pea varieties grown in different settings. Each bar represents the mean of five individual seeds. Absent bars indicate that no seed batch from this source was available. Error bars denote standard error. ** denotes significance at the p < 0.01 level. Individual seed measurements for the seed batch marked with a black arrow are displayed in Figure 13.*

The variation in seed ODAP concentrations observed between different growth conditions indicated that the seeds of accessions grown in different environments cannot be compared directly. This represents a limitation of the germplasm screening dataset, because differences between the environmental conditions between the JIC field in the UK (where the seeds of the IPK and USDA populations were grown) and the DZRC field (where the seeds of the EIAR population were grown) may have affected ODAP- concentrations. Because accessions may display genotype-specific responses to environmental conditions it

would not be permissible to test a small number of accessions in different environments and use these data to normalise ODAP concentrations of the entire population. For this reason, accessions in the EIAR population were not directly compared to the accessions in the IPK and USDA collections.

In addition, some batches showed very high seed-to-seed variations in ODAP concentrations (example shown in Figure 13). In this batch of Ratan, two seeds would have been classed as low-toxin, while three would not. While some seed-to seed variation may be expected due to the phenotypic plasticity, this extreme variation, only observed in some seed batches, suggests that these batches may have been contaminated, either by physical admixture of seeds of another accession during harvest or by inadvertent outcrossing of the plants.
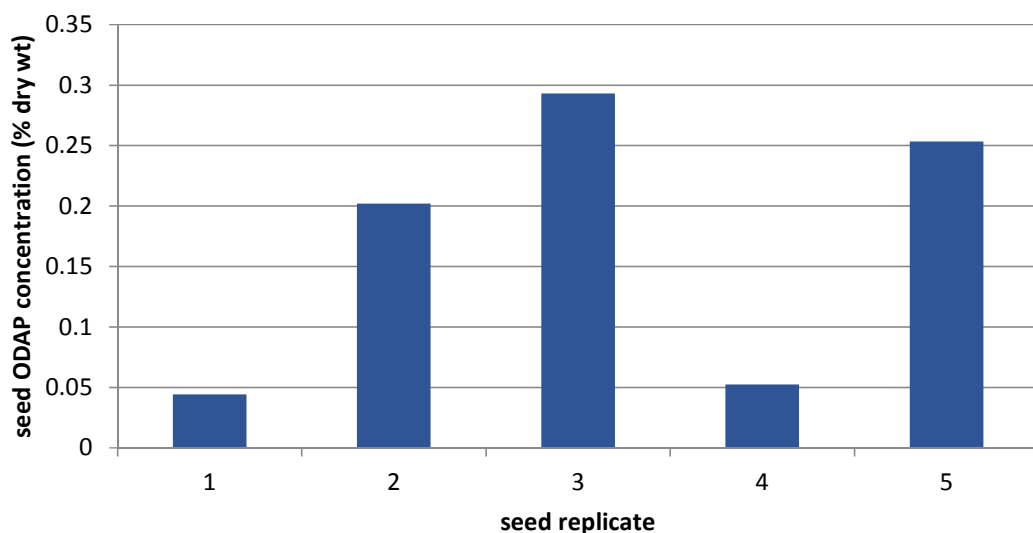
Figure 13. ODAP concentrations in individual seeds of Ratan (BioL-212) grown in a glasshouse at the JIC in summer 2015

Because the method of obtaining seed meal by drilling individual seeds is labour-intensive and cannot be easily scaled up, testing a representative sample of individual seeds to obtain accurate averages of seed ODAP concentrations of germplasm collections comprising hundreds of accessions would be prohibitively time-consuming. As an alternative, bulk samples of seed meal produced from dozens of individual seeds could be used to measure average seed ODAP concentrations of grass pea accessions. This would also result in a more meaningful measurement of an accession's seed ODAP content, as it

would be correlated directly with the absolute ODAP content of a portion of seeds to be consumed. To obtain the same average from testing individual seeds, a large number of seeds would have to be assayed to prevent the results from being skewed by accidentally selecting unrepresentative seeds. In addition the average calculated from individual seed measurements would have to be weighted according to the total mass of each seed measured to prevent small seeds with abnormal ODAP concentrations from distorting the average. Measuring individual seeds could also introduce issues of unconscious bias in selecting seeds for individual seed measurements and uneven distribution of ODAP inside the seed, as only the storage cotyledons were drilled during the collection of material from single seeds. For these reasons I decided to rely on bulk samples of seed meal for measuring average ODAP concentrations in the seeds of accessions from the germplasm collections.

2.3.3   ODAP concentrations in seeds of accessions from the IPK population

A population of 44 grass pea accessions collected from several countries cultivating grass pea by the IPK, Gatersleben, Germany, was assayed for seed ODAP concentrations using the methods described in sections 2.2.6 and 2.2.7. The ODAP concentration of each of the IPK population accessions as well as the Indian grass pea varieties grown and assayed alongside (using bulk samples) are shown in Figure 14. All the seeds used in this experiment were grown in the field at the JIC, Norwich, UK. The IPK population did not contain any accessions that would be classed as 'low-toxin' accessions based on the commonly used threshold of 0.1 % w/w. While there was significant variation between ODAP contents of some of these accessions, none of them stood out as breeding material for the development of new low-ODAP varieties. The ODAP concentrations of the accessions tested in this population did not correlate with their country of origin. For example, seed ODAP concentrations in the seven samples from Greece ranged from 0.259 % to 0.400 % w/w. The number and geographic range of accessions in this collection are limitations of this dataset. Important grass pea cultivating areas of the world, especially South Asia and the Ethiopian highlands are not well represented. Instead, the IPK collection contains many accessions from the Northern Mediterranean and Eastern Europe regions, where grass pea is cultivated only on very limited acreages and primarily as livestock feed (Milczak et al., 2001; Tavoletti et al., 2005; Başaran et al., 2011). The limited variation in ODAP levels in this population reflects this lack of diversity.
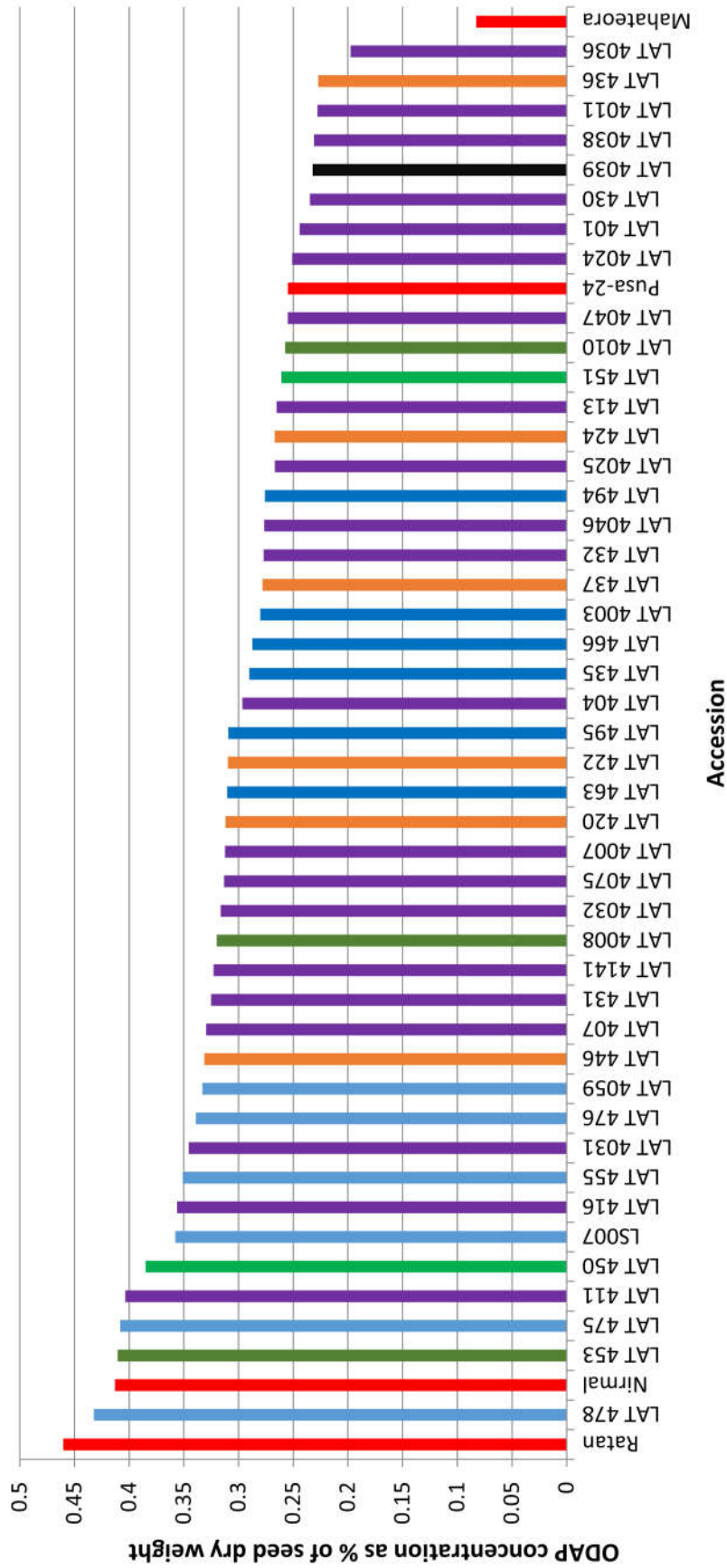
Figure 14. Seed ODAP concentrations of bulk samples of each accession of the IPK population and the Indian grass pea varieties, ordered by ODAP concentration. Measurements shown are based on three technical replicates. All seeds were grown in the field at the JIC in summer 2015. Colours indicate the geographic region of origin of accessions: India (red); Middle East (green); northern Mediterranean (purple); rest of Europe (blue); Africa (ochre); Peru (black); unknown origin (maroon). Countries of origin for each accession are given in Appendix 1.1.3.

The population exhibited 2.5-fold variation in seed ODAP concentration with a population mean of 0.31 % w/w with standard deviation 0.049 %. The distribution of ODAP concentrations in the population is shown in Figure 15. Seed ODAP concentrations and countries of origin for each accession are given in Appendix 1.3.
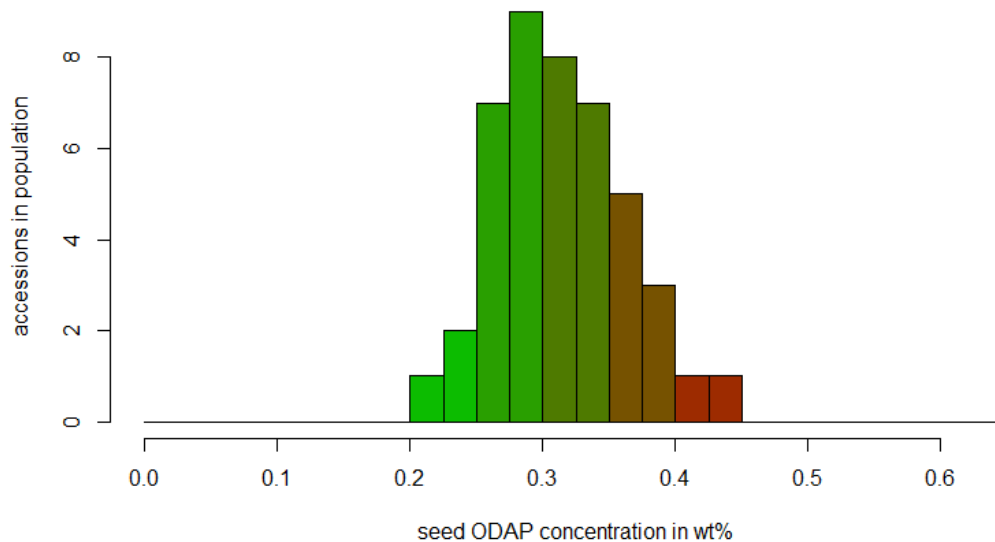
Figure 15. Distribution of seed ODAP concentrations of accessions from the IPK population. All seeds were grown in the field at the JIC, Norwich, UK in summer 2015.

2.3.4   ODAP concentrations in seeds of accessions from the USDA population

To increase the diversity of germplasm from grass pea cultivating regions for analysis of ODAP contents, I requested 96 accessions from the USDA seed bank at Pullman, Washington, USA. Inspection of the seed packets from the USDA population showed that the seed phenotypes within individual seed packets were highly diverse (example shown in Figure 16). In particular, many packets contained dark, speckled seeds as well as cream-coloured non-speckled seeds, along with variation in seed size. In total, out of 96 accessions requested from USDA, 48 contained two distinct seed types and 4 contained three distinct seed types. The proportions of seed types varied: in some accessions almost all seeds had the same seed morphology, while some accessions were equally split between their different seed types.

Figure 16. Photo of seeds of accession PI 255368 (Former Serbia and Montenegro) as received from the USDA seed collection. Seeds have been separated by seed morphology. Individual seed types were treated as separate sub-accessions in the following experiment.

This intra-accession morphological diversity could be caused either by phenotypic plasticity, leading to different appearances of genetically very similar seeds, or by an underlying genetic variability. The first appears unlikely, as this range of variation in seed morphology has not been reported for any genetically uniform accessions of grass pea, and did not appear in the other germplasm collections used in this study. If the phenotypic variation is caused by underlying genetic variation, this could be due to either a physical mixture of seeds, or a genetic mixture. A physical mixture would have resulted if seeds of genetically different mother plants were combined during harvesting or the later processing of the seeds. A genetic mixture would be present if the accession contained heterozygous individuals that caused traits to segregate when selfed. Both kinds of mixtures could reflect genetic diversity within the sample when the accession was first collected (accessions did not undergo single-seed descent) or be the result of accidental contamination or unwanted outcrossing during the regeneration of the accessions at the seed bank. If a simple physical mixture is present, sorting the seeds according to their morphology should resolve this, though it might not be clear which seed type represents the original accession. I therefore split up the seeds from highly variable seed packets into two or three groups as required and treated them as separate samples from then on. The seed morphologies and separation into subaccessions are shown in Appendix 1.4. If this variation exists due to heterozygosity in the original accessions, it does not pose a problem for a diversity screen. It is, however, possible that this heterozygosity within accessions in the germplasm collection has been caused by unwanted hybridisation events between different accessions in the several generations over which they have been maintained. Although the grass pea flower is cleistogamous and hence self-fertile, highly variable rates of outcrossing have been reported, ranging from as low as 2.2 % (Chowdhury and Sllnkard, 1997) to as high as

36 % (Gutiérrez-Marcos et al., 2006). If the accessions have been grown side-by-side in an open field without netting to prevent pollinators from accessing the flowers, it is possible that outcrossing has occurred.

When harvesting the seeds of plants grown in the field at the JIC (the offspring of the seeds obtained from USDA), it became apparent that many sub-accessions which I had separated based on their seed morphology produced seeds that again segregated for seed morphology (see Appendix 1.5). This suggested that the mixed seed morphologies observed in the previous generation were not just the result of physical contamination, but likely represented genetic mixtures resulting from outcrossing. This result is reminiscent of the large seed-to-seed variations in ODAP concentrations observed in the experiment comparing batches of seed of Indian grass pea varieties (see section 2.3.2). Outcrossing of breeding stock and accessions in seed banks could erode the diversity of stored germplasm collections and frustrate breeding efforts. It is therefore imperative that plants grown for the multiplication and regeneration of seeds of accessions in germplasm collections are kept under netting to prevent pollinators from accessing their flowers and providing outcrossing.

Seeds of accessions obtained from the USDA population were sown at the JIC and grown during the summer of 2015. Harvested seeds were ground up as bulk samples and assayed for their ODAP concentration in three technical replicates as described in sections 2.2.6 and 2.2.7[1]. The mean seed ODAP concentration of this population was 0.389 % w/w, with standard deviation of 0.123 %. The distribution of β-L-ODAP concentrations is shown in Figure 17. Seed ODAP concentrations and countries of origin for each accession are given in Appendix 1.3. The lowest-ODAP accession in the USDA population contained 0.040 % w/w ODAP. This was sub-accession A (separated according to seed morphology) of LS8246 (USDA accession ID PI 506418), a low-toxin variety developed in Canada (Campbell and Briggs, 1987) by selection breeding from the Indian variety Pusa-24. The seeds produced by sub-accession B of LS8246 had a much higher ODAP concentration (0.345 % w/w). This may indicate that sub-accession B is a physical contaminant of this low-ODAP accession. The sub-accession with the highest ODAP concentrationwas PI 667269/B USA IFLS 432 Sel 519 (0.763 % w/w). Sub-accession PI 667269/A also showed a high ODAP concentration (0.660 % w/w).

---

[1] The USDA population was assayed by Di Yang (DY) of the University of East Anglia (UEA) under my supervision. Experimental design was done by me and experimental work was done by DY. All data analysis presented in this thesis was performed by me. The data presented in this section, along with her own analysis, will formed part of DY's Masters thesis submitted to UEA.
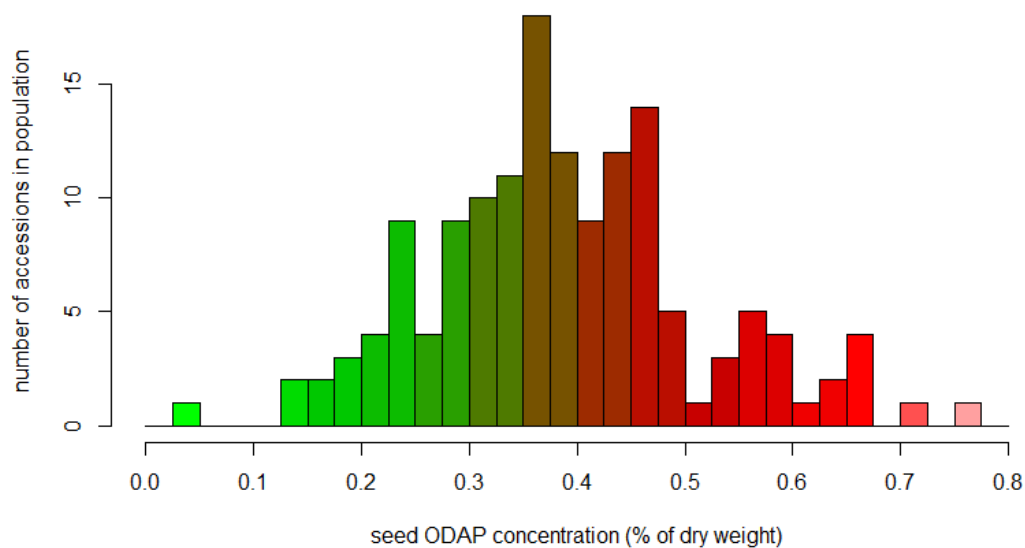
Figure 17. Histogram showing the distribution of ODAP concentrations in accessions in the USDA population. All seeds were grown in the field at the JIC in summer 2015.

Overall, the ODAP concentrations of accessions in the USDA population showed a much wider spread than the ODAP concentrations of accessions in the IPK population, which may reflect greater genetic diversity among the USDA population. No geographical pattern of toxin distribution was apparent among these samples. This population included breeding lines (from USA and India), local cultivars and landraces. Several accessions contained intermediate levels of β-L-ODAP (between 0.1 % and 0.2 % w/w). These were the two breeding lines from USA (IFLS 385 Sel 504 and IFLS 394 Sel 528), the Turkish cultivar Murdumuk, the Turkish landrace PI 206891 (sub-accession A) and the Greek accession CPI 14162 (sub-accession A). Among the Indian cultivars provided by BCKV, Pusa-24 and Nirmal showed intermediate ODAP concentrations. The cultivars LSWT11 and Mahateora, which were used as high- and low-ODAP controls in later experiments, were shown to contain 0.317 % w/w and 0.039 % w/w respectively. The cultivar Ratan was not included in this analysis as the evidence discussed in the previous section indicated that ODAP levels in this seed sample were highly variable, potentially due to contamination.

It is unlikely that differences in growth conditions at the seed centres at USDA, IPK and BCKV would have an impact on the β-L-ODAP levels observed in these experiments, because all the seeds used for the assays were grown in the field at JIC. An apparently transient epigenetic effect has been observed to affect β-L-ODAP concentrations in grass

pea tissues depending on environmental conditions. Shoot tips, stems and leaves of plants of the same genotype of grass pea, grown under well-watered conditions, contained higher concentrations of β-L-ODAP if the mother plant had been grown under simulated drought conditions. The difference between offspring of plants grown under simulated drought and well-watered conditions decreased over the lifetime of the offspring plants and seeds produced by these plants did not show significant differences in β-L-ODAP concentrations (Jiao et al., 2006). It is conceivable that similar epigenetic effects induced by the differences in the growth conditions of the previous generation could affect the β-L-ODAP levels in the plants grown at JIC. However, such differential effects would have likely worn off over the lifetime of the plants at JIC, allowing the seeds produced by these plants to be compared (Jiao et al., 2006)

2.3.5   ODAP concentrations in seeds of accessions from the EIAR population

The total ODAP concentrations in seeds of grass pea accessions from the EIAR population were measured using the spectrophotometric method described in sections 2.2.6 and 2.2.7 using three technical replicates of measurements based on one extraction from bulk seed meal samples. The population includes accessions obtained from Eritrea (3 accessions) and Australia (10 accessions), but most were collected from sites in Ethiopia (413 accessions). All seeds used in this experiment were grown during the same season at Debre Zeit Research Centre (Ada'a Chukala district, Oromia, Ethiopia)[2]. The distribution of ODAP levels in the population is shown in Figure 18. Seed ODAP concentrations and passport data for each accession are given in Appendix 1.3. The great majority of accessions fell within a distribution of ODAP levels with mean 0.368 % w/w and standard deviation of 0.051 % of dry seed weight. One accession, (ID 236701, originating from Bahir Dar in northern Ethiopia) contained a noticeably higher seed ODAP concentration of 0.61 % w/w. Eight accessions exhibited seed ODAP concentrations below 0.15 % w/w. Of these accessions, seven originated from Australia. These were produced by a breeding programme for low-β-L-ODAP genotypes and thus did not represent previously unknown low-ODAP germplasm. Seed ODAP concentrations of all the accessions from Australia are shown in detail in Figure 19.

---

[2] The EIAR population was assayed in collaboration with Alemu Abate (AA) of Aksum University, Tigray, Ethiopia. Experimental design was done by both of us, while most of the lab work carrying out the assay was done by AA, under my supervision. All data analysis presented in this thesis was performed by me. The data presented in this section, along with his own analysis, will also form part of AA's PhD thesis to be submitted to Aksum University.
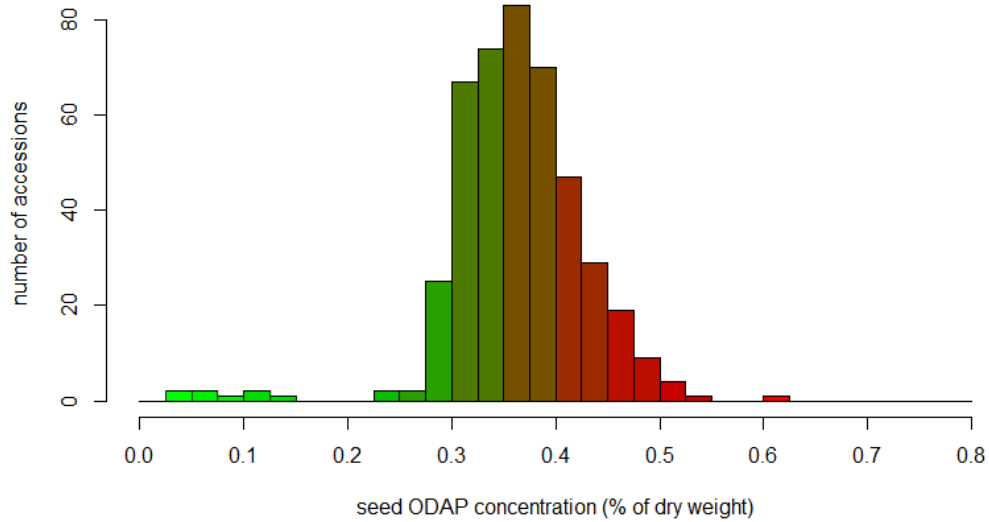
Figure 18. Distribution of seed ODAP concentrations in the EIAR grass pea germplasm collection. All seeds were grown at DZRC, Ada'a Chukala, Ethiopia
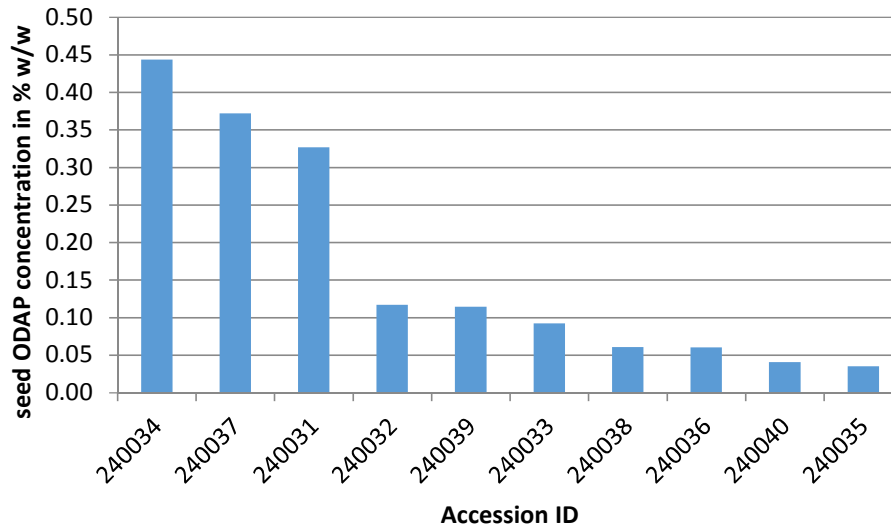
Figure 19. ODAP concentrations in seeds of grass pea accessions originating from Australia, ordered by ODAP concentration. Values shown are based on three technical replicates.

Of the accessions collected from Ethiopia, one (ID 242217 from the district Guba Lafto, North Wello zone, Amhara region) stood out for having a low seed ODAP concentration of 0.136 % w/w. It would not be classed as a low-toxin accession according to the 0.1 % w/w

threshold. However, this accession could be useful for the breeding of new low-toxin varieties if crossed with other accessions. This and the Australian low-ODAP accessions will be utilised in a pre-breeding programme for new low-ODAP varieties adapted to the Ethiopian climate that is now being planned by the EIAR.

The majority of accessions (395 out of 426) in the EIAR population have associated passport information on their geographic origin, including the administrative region and district (woreda) of origin. A smaller number of accessions (132) also have geographic coordinates detailing their exact origin, but these data are missing for most accessions in the population. The map depicted in Figure 20 shows the average seed ODAP concentration of the accessions from each district that is represented in the population. For comparison, false colour maps showing average annual rainfall and maximum temperatures are shown in Figure 21. The distributions of seed ODAP concentration in accessions from districts from which ten or more accessions have been collected are shown in Figure 22. No clear patterns of correlation between seed ODAP concentrations and collection sites emerge from these data. Accessions from the dry and hot Tigray region in the north of the country do not show higher or lower ODAP concentrations than accessions from the more temperate Amhara and Oromia regions in the centre of the country. The three westernmost districts represented in the population (districts Bure, Nejo and Nole Kaba, all in the Oromia region) all had seed ODAP concentrations below the population average. This is also the part of the country with the highest annual rainfall. However, as each of these three districts is represented by only a single accession in the population, this may not be a valid correlation. More accessions from this region would need to be tested to ascertain if a robust correlation exists.
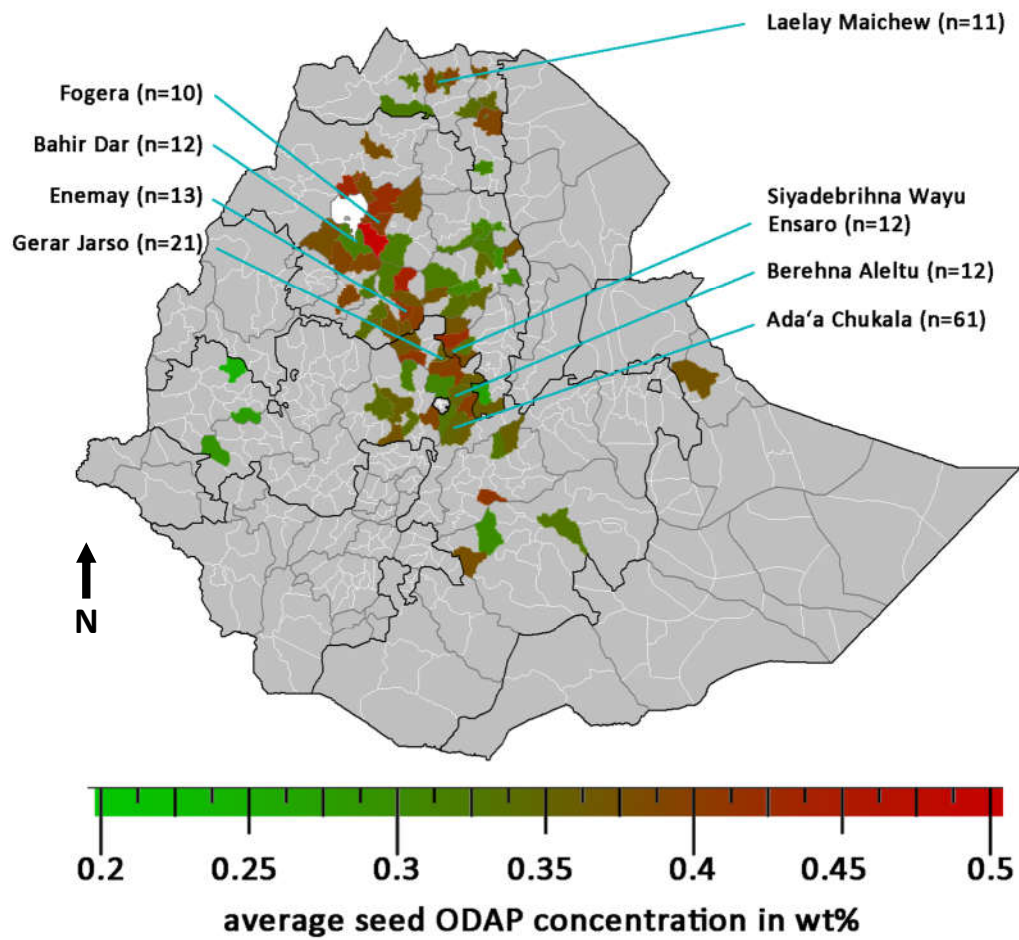
Figure 20. Map of Ethiopia, coloured according to the average ODAP concentrations of grass pea accessions originating from each district (woreda). Districts which are represented by ten or more accessions in the EIAR population are marked with the number of accessions sampled from there. Distributions of seed ODAP concentrations of accessions from these districts are shown in Figure 22. Figure produced using Photoshop CS5.

A

B
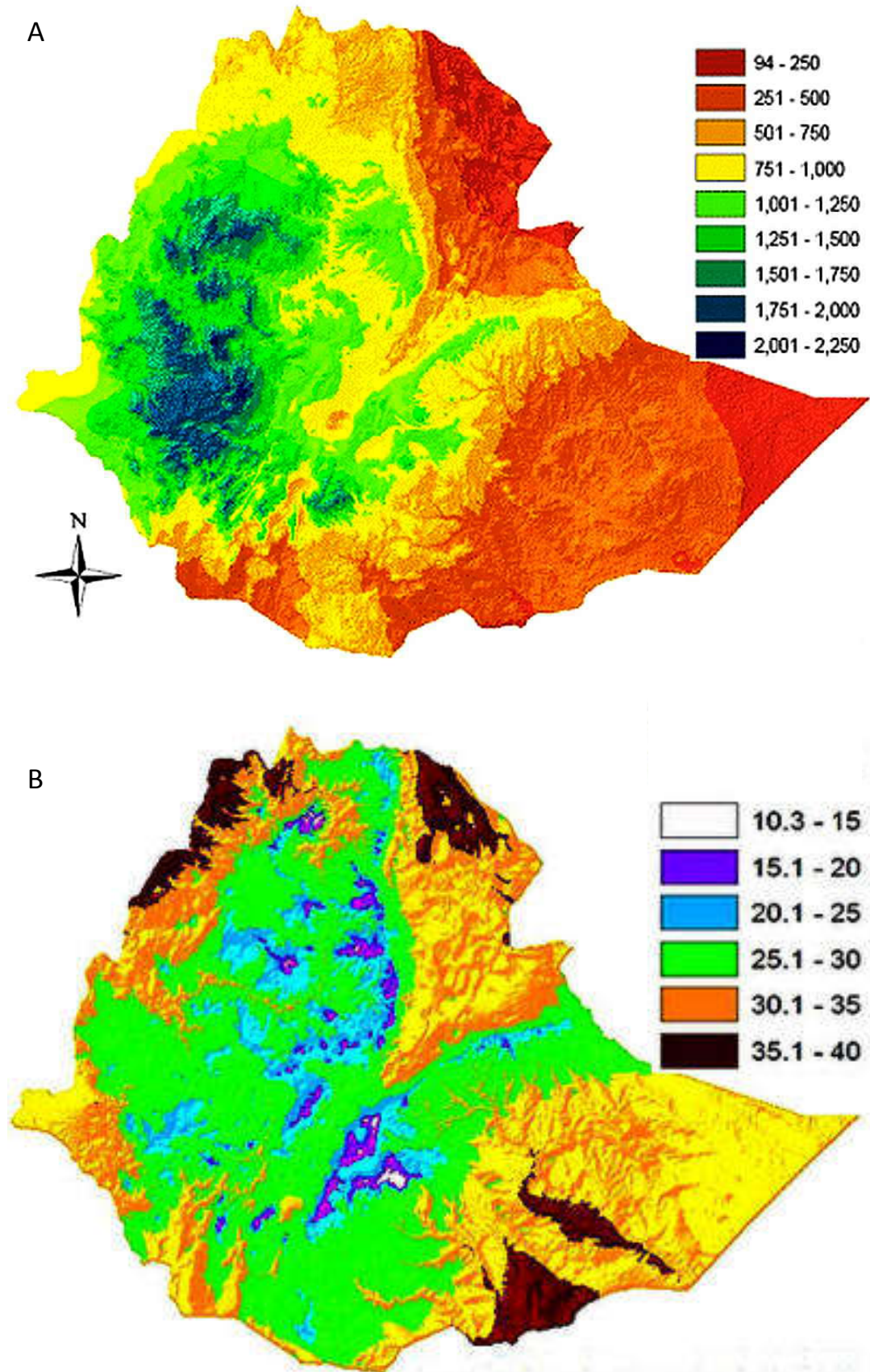
Figure 21. Maps of Ethiopia showing A) average annual rainfall in mm and B) average annual maximum temperatures in °C (Vreugdenhil et al., 2012). Figure reproduced with permission.
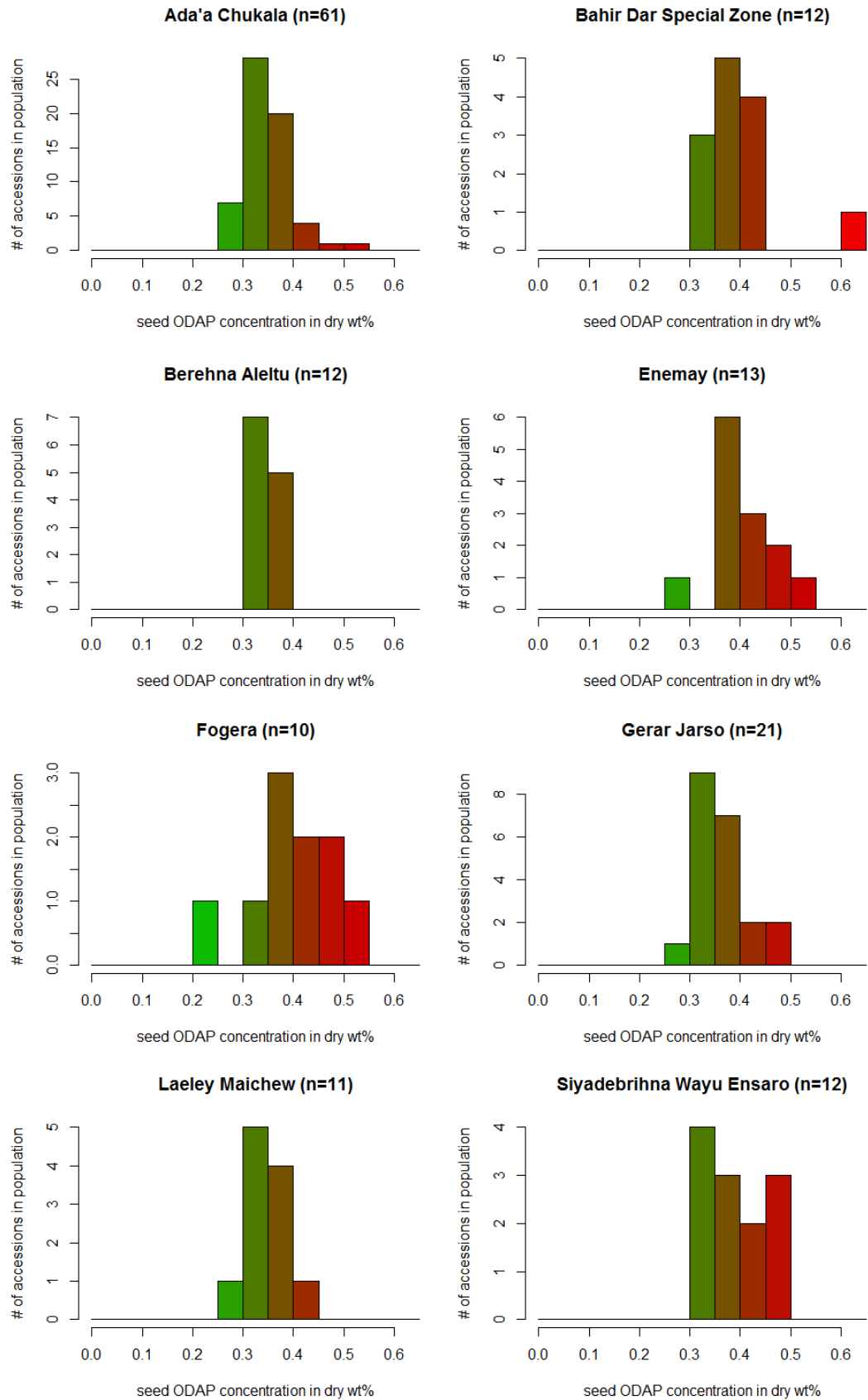
Figure 22. Distributions of seed ODAP concentrations of accessions originating from individual districts (woredas) in the EIAR grass pea population. Districts represented by ten or more accessions in the population are displayed.

For a subset of the EIAR population (287 accessions) the accession passport data included the altitude of the collection sites. The Ethiopian highlands are a geologically young region with an extremely varied landscape marked by steep sloped mountains and deep valleys (Williams et al., 2004). I plotted the seed ODAP conccentration of accessions for which altitude data were available against the altitude of their collection sites to test whether there was any correlation between the two factors (see Figure 23). All seeds used in this experiment were grown at DZRC, Bischoftu, Oromia, Ethiopia, at an elevation of 1920 m.
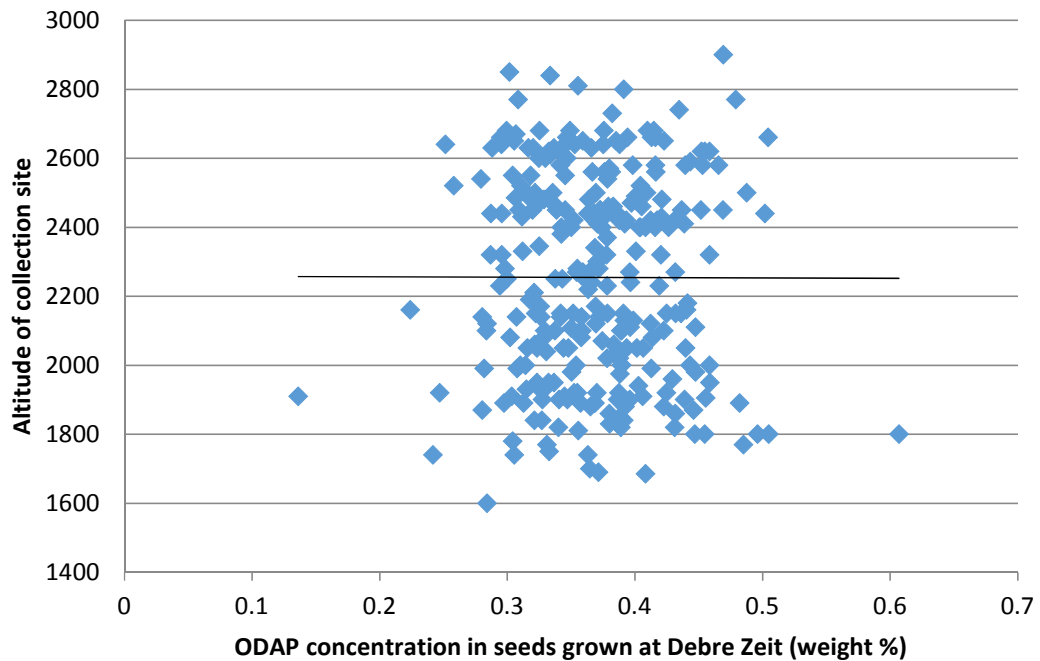
Figure 23. Seed ODAP concentrations of accessions of the EIAR population for which passport altitude data was also available plotted against the altitude of their collection site. A linear trend line is displayed.

No association between altitude of collection sites and seed ODAP concentration was observed. A linear regression produced a virtually flat slope (m = -3.6 x 10^{-7}) with extremely low explanatory power for the variation in the data (R^2 = 3.7 x 10^{-6}). This implied that there was no genotypic adaptation of the levels of seed ODAP to different altitudes, or at least none that persists when accessions are grown side by side. This could be because there is no evolutionary or breeding pressure selecting for different seed ODAP levels at different altitudes or because the amount of seed exchange and intercrossing of accessions between cultivation areas of different altitude obscures any genotypic differences in ODAP production that do exist.

The absence of associations between altitude or geographic origin and seed ODAP levels suggests that different levels of ODAP production do not confer advantages or disadvantages in the different agricultural settings in the geographic range investigated. In

contrast, the absence of any zero- or very low ODAP accessions in this landrace collection could imply that ODAP fulfils some physiological role in the plant and that its absence might be selected against by evolution or breeding. A comparison across a number of different growth conditions of near-isogenic lines of grass pea differing only in the ODAP production in their tissues would be necessary to ascertain whether low or zero-ODAP content in the seed would compromise the crop in agricultural settings.

2.3.6   Screening for L-2,3-diaminopropionic acid accumulating accessions

One limitation of any spectrophotometric assay based on absorbance at a single wavelength of visible light is a lack of specificity. Any other compound in the sample that also causes absorbance at 420 nm would distort the ODAP measurement. One way to exclude this background noise was to take a second measurement of absorbance at 420 nm of samples that had not been hydrolysed prior to mixing with the OPA/β-mercaptoethanol reagent buffer. Subtracting the absorbance of the non-hydrolysed sample after the colour-forming reaction removes the absorbance caused by the sample container, buffer, compounds unaffected by the hydrolysis and suspended particles. The difference between the two absorbance readings corresponds to compounds that were affected by the hydrolysis treatment, in particular ODAP, which is turned into L-DAP, and reacts with the OPA reagent.

Measuring the absorbance of the non-hydrolysed extract in the presence of the reagent buffer also accounts for any L-DAP that may already be present in the sample before hydrolysis. While L-DAP has never been shown to be present at measurable levels in grass pea tissues, it has been proposed as the intermediate immediately preceding β-L-ODAP in its synthetic pathway (Ikegami et al., 1999; Kuo et al., 1994; Kuo and Lambein, 1991). Any such L-DAP would react with β-mercaptoethanol and o-phthalaldehyde in both the hydrolysed and non-hydrolysed samples. By subtracting the absorbance values of the latter from the former, this could be excluded, leaving the difference as a measure of hydrolysed ODAP. Among the accessions screened in these experiments, none showed abnormally high absorbance readings at 420 nm in samples that had not undergone the hydrolysis treatment. If any such samples had been observed, this would have pointed to these accessions being L-DAP accumulators, which could have provided insight into the biosynthesis of β-L-ODAP. However, no accessions accumulating levels of L-DAP that could be measured using the spectrophotometric assay were present among the accessions included in this study.

2.4   Summary

I screened three grass pea populations by measuring the amount of ODAP in bulk seed meal samples using a spectrophotometric assay, with the aim of identifying grass pea accessions with low amounts of ODAP in their seeds. I screened two international germplasm collections, which included accessions from all countries in which grass pea is currently cultivated or has been cultivated in the recent past. I also screened a population of landraces collected from the Ethiopian highlands, one of the regions where grass pea is currently grown as a significant food crop. By including germplasm from a wide geographical distribution as well as germplasm that had been developed through centuries of informal selection in a highly divided landscape, I aimed to capture as much as possible of the existing genetic diversity of grass pea in order to identify low-ODAP genotypes that could be used for future breeding programmes. However, my screen did not identify any new low-toxin accessions – the only accessions with seed ODAP concentrations below 0.1 % w/w derived from breeding programmes and do not represent previously unknown low-toxin genotypes. Among the Ethiopian germplasm, one accession showed intermediate seed ODAP concentration of around 0.14 % w/w, which is similar to previously identified medium-toxin varieties such as Pusa-24 or Nirmal. This accession may be of value to breeders if ODAP levels are not greatly increased under conditions of environmental stress. No geographical patterns of seed ODAP concentrations were identified in either the international or the Ethiopian populations. These results reflect the difficulty that breeders have faced in developing low-ODAP grass pea material over the course of the past fifty years. As the natural variation for this particular trait appears to be limited, with only very few low-ODAP and no zero-ODAP accessions identified, it may be very difficult to breed zero-ODAP varieties using natural variation alone. I concluded that a means to induce greater genetic diversity for this trait might be necessary to achieve this breeding objective.

Chapter 3 – Identification of low-ODAP grass peas from a mutagenised population

3.1 Introduction

3.1.1   Increasing genetic variation by mutagenesis

The genes underpinning the synthesis of ODAP in plant tissues are still unknown and the available genomic and genetic resources for grass pea are limited. A classical approach to increasing the available variability of traits in plant populations has been to induce a higher than natural rate of mutations, by the application of physical, chemical or biological mutagens. The Mutant Variety Database maintained by the UN Food and Agriculture Organisation, includes more than 3200 varieties derived by mutagenesis of over 215 plant species, three quarters of them crops (FAO/IAEA, 2016) This is particularly useful in situations where the loss of a function, such as the biosynthesis of a toxin, is the goal. In the past, this approach has, for example, been used to reduce the concentration of antinutritional factors in common bean (Sparvoli et al., 2016) and soybean (Manjaya, 2009). Crucially, this kind of mutagenesis does not require any of the regulatory or biosynthetic genes to be known, as the screen operates entirely on the level of phenotypes.

In addition, mutagenesis as a tool for crop improvement is largely free of the legal regulatory hurdles that complicate the introduction of genetically modified crops in most jurisdictions (Parry et al., 2009). With the exception of Canada, currently the only country where mutant varieties are regulated as 'plants with novel traits' along with genetically modified varieties and some conventionally bred varieties (Jones, 2015), most jurisdictions do not place plant varieties derived by random mutagenesis under any specific regulation beyond rules applying to all new plant varieties (Kumar et al., 2017).

As there is no readily available means of eliminating the production of β-L-ODAP by conventional breeding (Kumar et al., 2011a), due to the absence of β-L-ODAP-free relatives in the gene-pool, a mutagenesis approach may be the most promising route to toxin-free grass peas. Drawbacks of mutagenesis are the high cost and practical difficulties of growing and screening a sufficiently large population of mutants, and the difficulty of identifying the causative mutations at the molecular level. To conduct a mutant screen for low-ODAP grass

pea mutants two things are needed: a mutant population of sufficient size and mutation density and a robust and efficient method of measuring ODAP.

3.1.2   Approaches to mutagenesis

A number of different strategies have been used to generate mutagenised populations of plants for research and breeding purposes. Mutagens broadly fall into three separate categories: physical, biological and chemical. Physical mutagenesis is achieved by bombarding plant tissues with high doses of radiation. The most commonly used physical mutagens are fast neutrons and γ-radiation. These physical mutagens disrupt the DNA, causing large deletions, which normally lead to the complete inactivity of affected genes. Biological mutagens are nucleic acids, such as transposons or T-DNA introduced using *Agrobacterium tumefaciens,* which are embedded in the genome often with a bias towards insertion in gene-rich regions. These typically cause the genes they insert into or near to to be inactivated, but they may also result in gain-of-function alleles, e.g. by activating a normally silent host gene. A drawback of physical and biological mutagens for use in a mutant screen is that they cause high mortality among mutants, limiting the density of mutations that can be achieved (Maple and Møller, 2007). Chemical mutagens are compounds that interfere with a cell's ability to repair its own DNA or themselves cause mutations by intercalating into the DNA double helix, causing errors in endogenous DNA replication. Chemical mutagens generally cause smaller scale mutations, such as single nucleotide replacements or small insertion/deletion mutations. These can result in the complete inactivation of a gene, but also in different levels of expression or activity of the gene product and, in rare cases, loss or gain of specific regulatory or metabolic functions of a gene. This can be an advantage compared to physical and biological mutagens, as it allows the generation of allelic series of mutants with varying levels of activity, in addition to null alleles. For example, the allelic series comprising a complete knockout mutation and six missense changes with less severe effects in the rice *OsDREB* (dehydration responsive element binding) gene provided promising material for the development of rice cultivars with improved drought resistance (Till et al., 2007).

Grass pea is a diploid (2n=14) species (Campbell, 1997). This limits the doses of mutagens that can be applied, as many essential genes may have only one copy in the haploid genome, causing any inactivating mutations to be homozygous lethal, resulting in a high death rate of mutants. On the other hand, the same reasoning makes it easier to discover

loss-of-function mutations of non-essential genes. In a series of studies published in 1976/77 Nerkar subjected several grass pea cultivars to γ-rays, as well as ethylmethanesulphonate (EMS) and N-nitroso-N-methyl urea (NMU), two commonly used chemical mutagens, finding EMS and NMU to be more effective at causing mutations, although the effects differed according to the grass pea genotype used (Nerkar, 1976, 1977b, a). These three mutagens were also compared in a 1987 study by Singh and Chaturvedi, who found γ-rays to be most efficient in causing mutations (i.e. causing the least damage per mutation), while being the least effective (in terms of mutations per comparable dose). NMU was found to be the most effective, but least efficient, while EMS was intermediate in both measures (Singh and Chaturvedi, 1987). These results contrast with the findings of a 2001 study by Waghmare and Mehra, who found EMS to be more efficient and effective than γ-rays (Waghmare and Mehra, 2001), as observed in other legumes (Kharkwal, 1998; Shah et al., 2008). A 2012 study by Tripathy et al. compared EMS and γ-rays to the chemical mutagen N-ethyl-N-nitro-N-nitrosoguanidine (ENNG) and found the latter to be the most effective and efficient mutagen (Tripathy et al., 2012).

Mutagenesis has already been used to increase the amount of available variation of traits in grass pea. The first attempt to generate low-toxin mutants by means of chemical and physical mutagens was made by Nerkar (1972), who claimed to have identified low-toxin mutant families, although no toxin-free mutants were found. However, a total of only 974 individual M2 plants (split into four separate mutagenesis treatments) were screened and the limited amount of detail given on experimental procedures and characteristics of the reported mutants restrict the usefulness of this study. A larger mutant screen using N-nitroso-N-methylurea (NMU) and sodium azide (NaN3), involving 1640 M2 families derived from the two Polish cultivars Derek and Krab, was conducted in 2003 (Rybiński, 2003). This study selected 20 mutant families from the population and then screened for variation in a range of phenotypes of leaf, stem, stipule, flower, pod and seed morphology, plant habit, branching and maturation as well as variation in agronomically relevant traits impacting on yield, including seed weight per plant, 100-seed weight and days to maturity. In a follow-up study, the researchers also identified mutants in seed microstructure (Rybinski et al., 2006). The researchers did not, however, screen for variation in ODAP content. As part of another mutant screen for morphological mutations, performed using gamma irradiation, dwarf mutants, distichous pedicel (double-flowered) and tendril-less mutants of grass pea have been identified (Talukdar, 2009a, b, 2013) Despite the obvious utility of toxin-free grass pea varieties and the likely amenability of this breeding goal to a mutant screen approach, no

sufficiently large screen has been published identifying low- or zero-toxin mutants. This is likely due to the low priority of grass pea for most research funding bodies and the practical difficulty of screening a large population of grass pea mutants for their ODAP content.

3.1.3   Selecting an appropriate screening method

As discussed in the previous chapter (see section 2.1.4), a number of methods for measuring β-L-ODAP in grass pea and ginseng (*Panax spp.*) tissues have been described in the literature, including capillary zone electrophoresis (CZE) (Arentoft and Greirson, 1995), high performance liquid chromatography (HPLC) (Yan et al., 2005), liquid chromatography mass spectrometry (LCMS) (Koh et al., 2005) and gas chromatography mass spectrometry (GCMS) (Xie et al., 2007). The method that has been used most widely to measure ODAP in grass pea samples, especially when large numbers of samples were processed, is the spectrophotometric method, also called the 'Rao-method' (Rao, 1978), which has been modified and improved by other researchers (Briggs et al., 1983; Hussain et al., 1994). Although it is less accurate and sensitive than some other methods, its low cost in terms of equipment and consumables have made it the method of choice for many laboratories (Tadesse and Bekele, 2003b; Campbell, 1997; Srivastava and Srivastava, 2006). Crucially, the chemistry of this assay method allows the processing of many samples in parallel, which can be measured rapidly using a plate spectrophotometer. In contrast, CZE and chromatography-based methods all rely on sequential measurement of samples. These processes can be automated using an autosampler, but not parallelised without prohibitive costs. While these methods are clearly preferable for the accurate measurement of ODAP content in a small number of samples, they cannot easily be scaled up to assay the tens of thousands of samples that need to be processed during a mutant screen. The aim of my mutant screen was to identify mutants with severely reduced ODAP content and in particular entirely ODAP-free mutants. Because of the large expected difference between wild-type-levels of ODAP and the ODAP contents in mutants in ODAP synthesis, a highly accurate assay was not required. Hence, the level of accuracy provided by the spectrophotometric method was sufficient for this application. After developing a scaled-down, plate-based version of the assay, which I used to measure ODAP concentrations in the seeds of grass pea germplasm collections, as described in the previous chapter, I improved the assay for the purposes of a mutant screen by parallelising and streamlining

several of its steps. The assay cannot differentiate between the two isomers α- and β-L-ODAP, but instead measures total ODAP levels. Hence the mutant screen could strictly speaking only identify low-ODAP mutants, rather than low-β-L-ODAP mutants, but because β-L-ODAP represents 95 % of the total ODAP in grass pea tissues (Roy and Rao, 1968) and the two isomers are able to interconvert in aqueous solutions (Abegaz et al., 1993), a measurement of total ODAP was considered sufficient to identify low-β-L-ODAP mutants.

3.1.4   Plant tissue used for the mutant screen and adaptations of the screening method

The ODAP toxin is present in all tissues of the plant, but accumulates to the highest levels in seeds and young tissues of the shoot. ODAP concentrations are particularly high in young shoot tips of seedlings, in young leaves and immature seeds (Jiao et al., 2006; Srivastava and Srivastava, 2006). Flowers, roots, stems, tendrils and mature leaves of the plant contain only small amounts of the toxin. It is unclear whether the toxin is synthesised in all of these tissues or is synthesised only in some tissues and then transported into others. As the toxin is a very small water-soluble metabolite, it is conceivable that it might be passively transported into some tissues, without a specific transport mechanism. The spectrophotometric assay for ODAP was first described for seed meal samples because this is the tissue most used for human consumption. It has since been used to measure ODAP concentrations in other tissues as well. However the use of seed meal samples poses two problems for a large-scale mutant screen.

Firstly, it is conceivable that ODAP may be synthesised in maternal tissues and deposited into the seed during seed development. Thus, even if an embryo of the M2 generation carries a homozygous mutation that prevents all ODAP production, its cotyledons may contain some ODAP that has been synthesised in its heterozygous mother plant. In order to prevent this, the next generation of seeds (M3) would have to be assessed, which would require many thousands of M2 plants to be grown to maturity.

Secondly, grass pea seeds are very hard when dry and do not break when shaken with grinding beads. The only ways to obtain seed meal is by using a drill or mill. Drilling seeds is a useful method to obtain meal from the storage cotyledons of a seed without destroying the embryo, allowing the same plant to be planted and grown to maturity. However, it is a laborious procedure, as each seed has to be individually chipped to remove the tough testa, fixed under the drill and carefully drilled to collect the meal. Afterwards, the drill

needs to be cleaned well to prevent contamination. While mills or grinders can break down the testa, they require larger amounts of seeds, necessitating the bulking of seeds (which might be segregating for ODAP content). This may lead to mutants in segregating M2 families being overlooked and only mutant families with the necessary number of seeds can be included. In addition, cleaning out a mill after each sample is even more time consuming than cleaning a drill.

An alternative to using seed meal samples is to collect samples of young shoots that have just emerged from the soil. This tissue contains very high concentrations of ODAP, even exceeding levels in the seeds (Jiao et al., 2006). As this is a tissue grown independently from the mother plant, the risk of confusing maternal effects is reduced. However, it is still possible that part of the ODAP measured in the young shoot has been synthesised in the mother plant, deposited in the seed and then translocated into the young shoot upon germination.

Shoots contain chlorophyll, which absorbs light at the wavelength that is used in the spectrophotometric assay for ODAP (420nm). If chlorophyll is extracted alongside ODAP, this may result in an unwanted background reading in the assay. For this reason, water needs to be used for the extraction, rather than ethanol, which has been used in some variations of the Rao method (Briggs et al., 1983), as the solubility of chlorophyll in water is much lower than in ethanol.

Due to the number of samples that need to be collected, variation in the time needed to imbibe and germinate seeds and unavoidable variation in the orientation of the seed and the planting depth, the stage of development of the seedlings varies after seven days. Instead of collecting samples from each mutant family after the same amount of time, samples should be collected at the same stage of seedling development, as far as possible.

Due to variability resulting from collecting shoot tips by hand from seedlings at varying stages of development, some variation in the fresh weight of collected samples was unavoidable. It might have been possible to weigh each sample after collection in order to normalise the results of the assay by the weight of each sample. However, this slows down sample collection significantly. Sample weight variation was estimated to be unlikely to account for more than a twofold variation in the assay result based on a preliminary experiment weighing a subset of the selected shoot tip samples (data not shown). The aim of the mutant screen was to find metabolic mutations that make the plants incapable of producing the toxin. Thus the only reading expected in true zero-ODAP mutants would be

at the background level due to the limited sensitivity of the assay and potential maternal effects. Therefore, I chose not to weigh samples before extraction, and instead rely on retesting to remove false positives that were selected only because of their low sample weight.

3.1.5   Pathway analysis by mutant screening

Mutants found by screening a mutagenised population can be used to study biochemical pathways (Huang and Sternberg, 2006). In many cases, the phenotype that the mutant screen selects for, such as the abundance of a specific metabolite, is affected by several genes. These include genes that encode enzymes, which catalyse successive steps in the metabolic pathway and regulatory genes, which encode gene products that regulate the expression and activity of these enzymes. By crossing mutants and testing the F1 generation for whether they exhibit a mutant or wild type phenotype for the trait in question, it is possible to identify complementation groups, each of which is composed of individuals carrying different mutations in the same gene. By screening enough mutant families and assigning them to complementation groups, it is possible to ascertain the number of metabolic and regulatory genes that significantly affect the trait. By analysing the metabolite profiles of these mutants it is possible to identify the metabolic steps that are affected by the mutations in each complementation group. Metabolites that accumulate beyond the wild type levels are likely to be upstream of the enzyme affected by the mutation, while metabolites with reduced levels in the mutant are likely to be downstream of the affected enzyme. These data can also be used to support or refute whether certain metabolites are intermediates in a biochemical pathway.

3.2   Materials and methods

3.2.1   Origin of the mutant population

The mutant population used in this screen was provided by Manash Chatterjee of BenchBio, Gujarat, India. The parent variety used in the mutant screen, designated LSWT11, was obtained from an Indian plant breeder. The variety is a common landrace in this part of India, with a medium to high ODAP content, small, dark and mottled seeds and blue flowers. The variety did not undergo single seed descent before being used for mutagenesis. Mutagenesis was performed by the company BenchBio using a proprietary mutagenesis protocol (Chantreau et al., 2013) employing EMS at a concentration of 0.75 %. This method of mutagenesis was used to generate a mutant population that could also be used for a reverse genetic TILLING (targeting induced local lesions in genomes) approach.

The mutagenized generation of seeds (M1) was sown in India and grown to maturity to collect M2 seeds. Each M1 plant was grown in isolation to prevent outcrossing. All seeds from one M1 plant formed one M2 family and each M2 family was collected into an individual seed packet. All seeds were treated with the fungicide Thiram to prevent growth of mould. An overview of the process used to generate the mutant population and of the terminology used to describe the different generations of seeds and plants is shown in Figure 24.

A   M1 seed – treated with 0.75% v/v EMS solution to cause mutations

B   M1 plants – grown in isolation, each giving rise to one M2 family by self-fertilisation

C   M2 seed were treated with Thiram to prevent moulding and collected into seed packets (one packet per family)

D   M2 seedlings were used for the ODAP assay, the plants were left to re-sprout after sample collection

12x   12x   12x

E   selected M2 plants (low-ODAP and morphological mutants) were potted up and grown to maturity

other M2 seedlings were discarded

F   M3 seeds were collected and used for further experiments

Figure 24: Schematic representation of the procedure used to generate the material used in the mutant screen and of the terminology used in this thesis. Steps A, B and C were performed by BenchBio, Vapi, Gujarat, India, steps D, E and F were performed at JIC as part of this thesis.

3.2.2   Assessment of mutation density

To estimate the mutation density of the mutant populations, large-effect chlorophyll mutations, such as *albino*, *xantha* and *chlorina* mutations were counted among the seedlings of the mutant population while conducting the screen. In addition, DNA was extracted from one plant each of 768 M2 families to measure heterozygosity and mutation density. Leaf samples were collected into 1.2 ml sample collection tube strips, which were stacked into plates for parallel processing. Four young leaflets were collected from each plant, aiming for 100 mg of fresh weight. Samples were flash frozen in liquid N_2 and lyophilised in a Virtis BenchTop SLC® freeze-drier (Virtis, Gardiner, New York, USA) for 48 h. A 4 mm steel grinding ball was added to each sample. Samples were ground by shaking at 20 Hz for one minute, followed by inverting the tubes and shaking at 20 Hz for another minute. Tubes were centrifuged at 2500 g for 20 seconds to detach sample powder from the lids of the collection tubes. To extract DNA, 750 µl of extraction buffer (200 mM Tris-HCl pH 7.5, 2 M NaCl, 25 mM EDTA, 0.5 % v/v sodium dodecylsulfate (SDS), 2 % v/v polyvinylpyrrolidone (PVP)), pre-heated to 65 °C, were added to each sample. Samples were further homogenised and re-suspended by shaking for 1 minute at 30 Hz, inverting the plate and shaking again. Plates were clamped between metal plates to prevent tubes from opening and submerged in a water bath at 65 °C for 60 minutes, followed by centrifugation at 5000 g for 15 minutes. The supernatant from each tube was transferred into a new tube and 400 µl of phenol:chloroform were added. Tubes were sealed immediately, inverted ten times to mix the contents and centrifuged at 5000 g for 15 minutes. The upper (aqueous) phase was removed to new collection tubes and an equal amount of chloroform was added. Tubes were again sealed, inverted ten times and centrifuged at 5000 g for 15 minutes. The aqueous phase was transferred to new collection tubes. DNase-free RNase was added to a concentration of 20 µg/ml followed by incubation at 37 °C for 15 minutes to degrade contaminating RNA. To precipitate DNA, 40 µl 3M sodium acetate pH 5.2 and 400 µl 100 % isopropanol were added and mixed by inverting the tubes. Tubes were left at room temperature for an hour and centrifuged at 5000 g for 45 minutes. The supernatant was poured off and the pellet was rinsed with 400 µl 70 % v/v ethanol (pre-chilled on ice) and left at 4 °C for 1 hour. Samples were centrifuged at 5000 g for 10 minutes to reattach the pellets. The supernatant was poured off and the washing step (adding ethanol, incubation at 4°C and centrifugation) was repeated once. The supernatant was removed and pellets were left to air-dry. Dried DNA samples were shipped to Brad Till at the joint laboratory of the International Atomic Energy Agency (IAEA) and the

Food and Agriculture Organisation (FAO) of the United Nations in Seibersdorf, Austria. The heterozygosity present in the parent variety and the mutation density in the mutant population were estimated by Ayse Sen and Brad Till using the methods described by Till et al. (2006).

3.2.3   Optimisation of seed scarification

A number of scarification and seed priming techniques were compared to identify the most suitable method to improve germination rate and synchronicity for the mutant screen. Dry control seeds were compared to seeds that had been scraped using sandpaper and seeds that had been chipped using scissors. In addition, chipped and non-chipped seeds were soaked in water overnight. All seeds (12 seeds each of untreated and dry scraped sets, 24 seeds each of wetted only, wetted scraped and wetted chipped sets) were sown into individual pots containing John Innes No. 2 compost with 20 % v/v grit. Germination was scored 7 days later.

For sandpaper scarification, seeds were placed on a foam mat soft enough that the seeds could be slightly pressed into it to hold them in place during scratching. Because some seeds were liable to break during the scarification process, 15 seeds from each M2 family were scarified together, to ensure that at least 12 would be intact for sowing. Seeds were then scratched using medium-coarse sandpaper glued to the bottom of a pipette tip box as shown in Figure 25. The sandpaper was surrounded by a 2 mm deep plastic edge that prevented it from scraping directly on the foam mat. Holding the pipette box, it was possible to press the seeds into the foam to stop them from flying out and scarify them by vigorously moving the box back and forth for about 15 seconds.
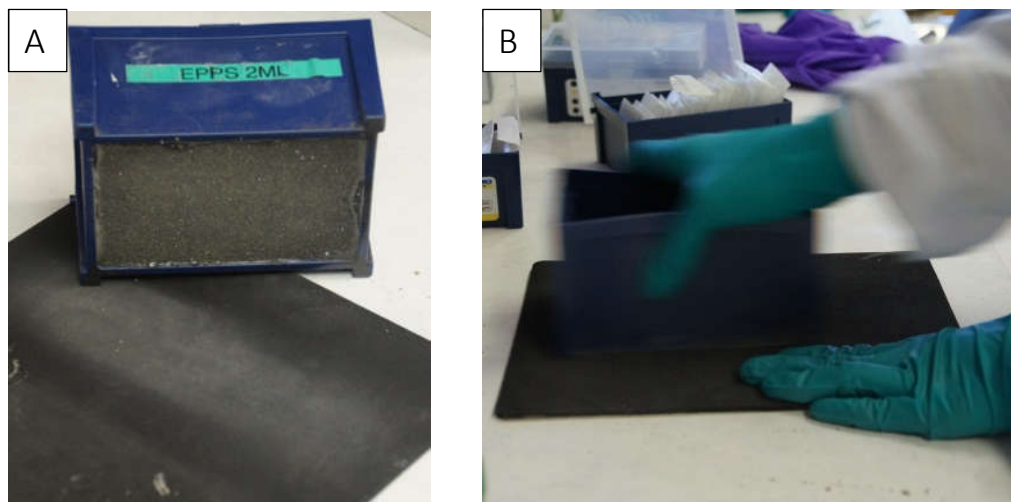
Figure 25. Rapid scarification method for grass pea seeds A) Sandpaper at bottom of pipette tip box and foam mat (face-down mouse mat) B) 15 seeds of each M2 family were rubbed for about 15 seconds.

3.2.4   Growth conditions and sample preparation for the mutant screen

The mutant screen was conducted in two parts: a pilot screen including 768 M2 families and a main screen including the remaining 2352 M2 families. The purpose of the pilot screen was to estimate the variability of ODAP content in the population, test all the methods associated with the mutant screen and resolve any problems. Consequently, minor changes were made to some of the methods for the main screen, as described in the following.

Seeds of M2 families were scarified with sandpaper as described in the previous section. The scarified seeds were sown individually into 13x22 well trays (pilot screen) or 12x19 well trays (main screen) filled with John Innes No. 2 compost with 20 % v/v grit, covered with sieved John Innes No. 2 and watered immediately. The trays were then placed in a lit (HQI lamps), temperature-controlled (18°C during night, 25°C during day) glasshouse for the pilot screen (September to December) or an unlit, heated glasshouse (min. 15°C) for the main screen (April to June), as shown in Figure 26. During the main screen, the trays were separated using white transparent perspex dividers, in order to prevent plants from different trays from getting entangled. Plants used in the crosses (October to February) were grown under sodium lamps in a heated glasshouse (min 15°C).
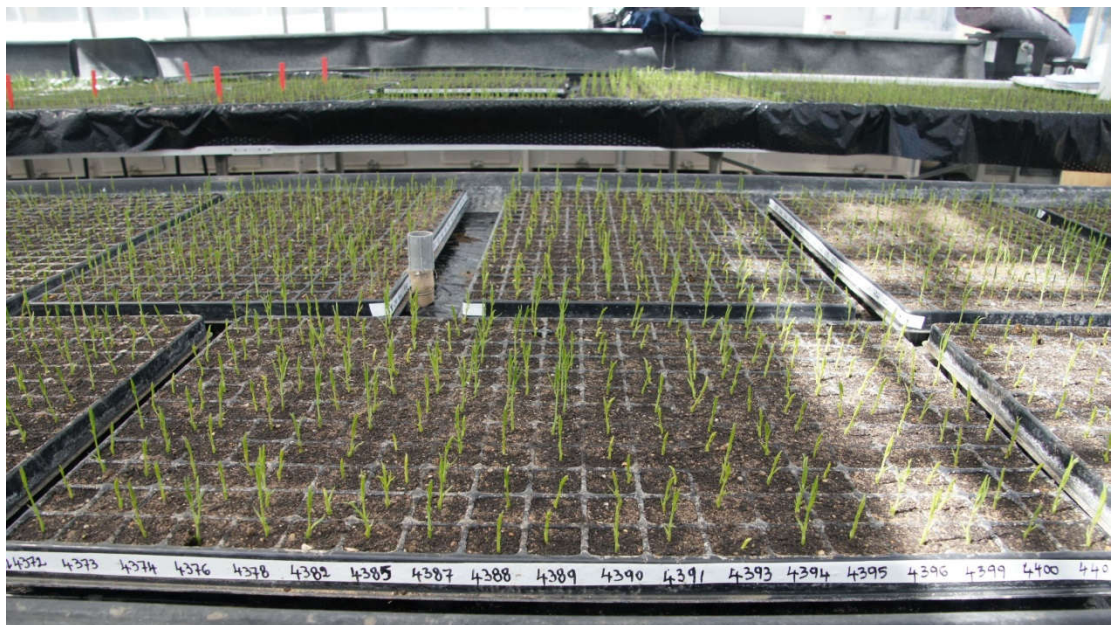
Figure 26. Arrangement of plants in 12x19 well trays during the main screen, just before harvesting. Each column of seedlings in a tray represents one M2 family.

For the pilot screen, three additional seeds of each family were scarified and planted in the thirteenth row at the bottom of each tray. These seedlings were then used to replace those seedlings in the other twelve rows that did not germinate, just before harvesting shoot tips. All seedlings in the thirteenth well that were not needed were discarded. While these replacement seedlings allowed for a reduction in the number of missing samples in the assay, it proved to be very laborious and slowed down the collection of samples significantly. This would have made it difficult to collect a much larger number of young shoot samples in one day, as was needed for the main screen. It was therefore decided not to sow replacement seeds in the main screen. This also made it practical to switch to larger trays with only 12 rows, which allowed more soil and space for each seed.

Seven to nine days after sowing, depending on the growth stage of the seedlings, cuttings between 5 mm to 10 mm long from the top of each seedling were collected, aiming for 15 mg of fresh weight as shown in Figure 27. Samples were collected using atraumatic surgical tweezers with horizontal teeth (Graefe Fixation Forceps. World Precision Instruments, Sarasota, Florida, USA), which allowed harvesting without crushing the samples, thus reducing the risk of cross-contamination. The samples were placed into 1.2 ml deep-well polystyrene plates which were sealed with a gas-permeable film. The plates were stored on dry ice prior to freeze-drying.

Figure 27. Sample collection using Graefe fixation forceps.

For each sample plate a form was completed to record which mutant family was placed in each row of the plate. As twelve samples were collected of each mutant family, each family was placed in a single row on the plate. Seeds that failed to germinate were skipped, resulting in empty wells on the collection plate. Skipped seeds were noted down on the collection form, so they would not result in false positives, i.e. samples that were counted as ODAP-free. The dates of sowing and sample collection as well as any unusual observations such as *albino* seedlings were also noted. Each collection form was then digitised using an html-form (see example form in Appendix 2.1), which created a text-file containing the metadata for each plate. This file was used by the data analysis script (described below) to match absorbance readings to the individual plants (listed by their positions in the growth trays) and to exclude missing samples.

As the assay procedure was more time consuming than the sampling procedure, some samples had to be stored until they could be assayed. Samples were lyophilised for 36 h in a BenchTop SLC® (Virtis, Gardiner, New York, USA) freeze-dryer. After this, the samples in the 96-well collection plates, were sealed in gas-impermeable bags together with a pouch of silica gel, and were stored at 4°C prior to the assay.

Plants were left to grow while assays were conducted and rapidly developed new branches from cotyledal buds or, where present, axillary buds. Leaf samples were collected from the same plants. Perspex dividers were placed between trays as shown in Figure 28. Plants that were selected as putative low-ODAP mutants were re-sampled by harvesting the top leaflets eight and ten weeks after sowing for the second and third passes of the mutant screen, respectively. At this stage, plants that had produced flowers were screened for flower colour mutations, which may eventually be useful as phenotypic markers for varietal development. All plants that had been selected in the third pass of the mutant screen and all flower colour mutants were transferred into individual pots. Other plants were discarded.

Figure 28. Three weeks after sowing, perspex dividers were inserted between the trays to stop plants from different trays from getting entangled.

3.2.5   The high-throughput spectrophotometric assay

To aid the extraction of ODAP, the samples had to be homogenised. A 4mm steel ball was added to each well containing a freeze-dried tissue sample (see Figure 29) and each plate was covered with a reusable silicone sealing mat.
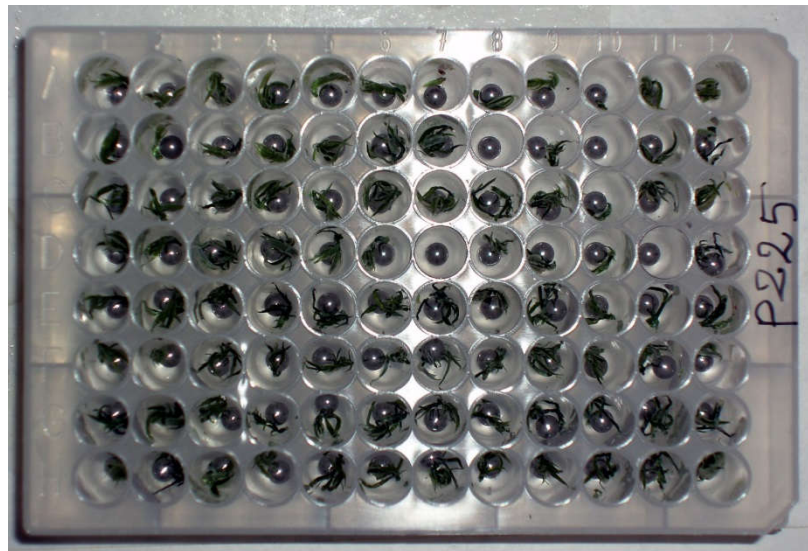
Figure 29. Deep-well plate containing freeze-dried shoot tip samples and steel balls prior to grinding

The samples were pulverised by shaking the plates for 30 seconds at 18 Hz, before inverting the plate and repeating. Shaking frequencies above 18 Hz had to be avoided as higher

speeds risked the bottoms of the extraction plates being shattered by the grinding balls if no liquid medium was present. Care was also taken to remove any glue debris, which could fall into the wells if plates were sealed with self-adhesive film and placed into a -80°C freezer, since glue in the wells prevented proper grinding.

To extract ODAP from ground tissue samples, 600 μl of sterile dH$_2$O was added to each sample and the reusable silicone sealing mat was replaced. Shaking was repeated for 30 seconds at 25 Hz for two orientations of the plate to suspend the sample meal and to break down the tissue further. As the water in the container softened the impact of the grinding bead, this shaking frequency did not break the bottom of the container. The plates were clamped between aluminium plates to prevent the build-up of pressure inside the wells from breaking the seal. The plates were then submerged in a water bath at 95°C for 20 minutes. This arrangement is shown in Figure 30.

After centrifugation of the plates for 10 minutes at 2500 g, 80 μl of extract were transferred into a PCR plate containing 160 μl of 3M KOH. The plates were sealed, clamped and submerged in the water bath at 95°C for 30 minutes.
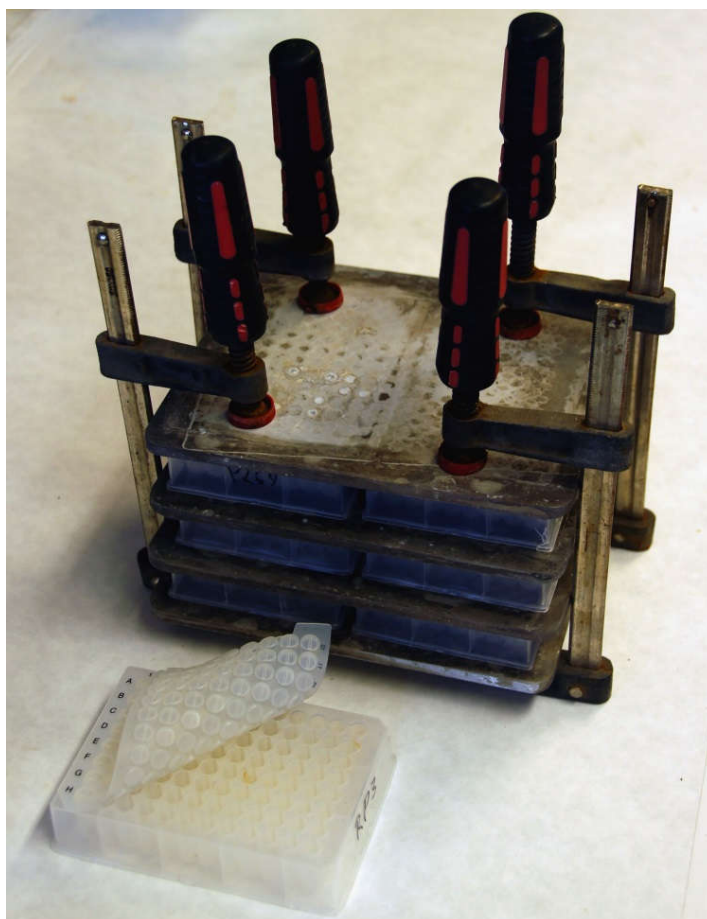
Figure 30. Stack of deep-well plates containing ground seedling shoot tip samples and water and sealed with silicone mats. Collection plates are separated by aluminium plates and held together by clamps

To prepare 100 ml of the o-phthalaldehyde/tetraborate reagent for assaying ODAP, 82 mg of o-phthalaldehyde (Sigma-Aldrich, St. Louis, Missouri, USA) were dissolved in 818 µl ethanol and 164 µl β-mercaptoethanol. Separately, 12.5 g of potassium tetraborate tetrahydrate (Sigma-Aldrich, St. Louis, Missouri, USA) were dissolved in 99 ml of dH20, forming a 0.409 M solution. The mixture was gently warmed in a water bath to help dissolve the potassium tetraborate tetrahydrate and allowed to cool to room temperature before the OPA/ethanol/β-mercaptoethanol solution was added. The spectrophotometric method described by Briggs et al. uses a more concentrated version of this reagent buffer with a potassium tetraborate concentration of 0.5 M, but additional dH_2O is added to the hydrolysed sample extracts prior to the colour-forming reaction (Hussain et al., 1994; Briggs et al., 1983). The reagent could be kept at 4 °C for up to three days before use.

Two clear optical flat-bottomed plates (Greiner Bio One, Kremsmünster, Austria) were prepared for each plate of samples. In the first, 30 µl of hydrolysed extract and 220 µl of

OPA/tetraborate buffer were added to each well. In the second, 20 µl of 3 M KOH, 10 µl of non-hydrolysed extract and 220 µl of OPA/tetraborate buffer were added, and left for no longer than five minutes between combining the KOH solution and extract before addition of the tetraborate buffer. The absorbance at 420 nm was measured in each well of hydrolysed and non-hydrolysed sample. Readings were taken at 25°C using an automated absorbance plate reader (VersaMax, Molecular Devices, Wokingham, UK). The procedure of the spectrophotometric assay is summarised in Figure 31.

Collect 10 - 20 mg of seedling

shoot tips (fresh weight) on dry ice

Freeze-dry for 36h

Add a grinding bead to each well

Grind samples in plate using TissueLyser

Add 600 µl dH$_2$O to the sample and grind again

Incubate at 95 °C for 20 min in a water bath

Centrifuge at 2575 g for 10 min

prepare PCR plate with 160 µl of 3 M KOH per well

Add 80 µl of extract to each well

Incubate plate at 95 °C for 30 min

in a water bath

Transfer 30 µl of hydrolysed

extract to a flat-bottom plate

add 220 µl OPA/tetraborate buffer

Shake and incubate at

room temp. for 15 min

Measure absorbance at 420 nm

using plate reader

Transfer 10 µl of non-hydrolysed

extract to a flat-bottom plate

add 20 µl of 3 M KOH and 220 µl

OPA/tetraborate buffer

Shake and incubate at

room temp. for 15 min

Measure absorbance at 420 nm

using plate reader

Figure 31. Overview flowchart of the spectrophotometric method for measuring ODAP in plate format

3.2.6   Automated data handling

Raw data from the assay were handled by a specially designed R-script (see Appendix 2.2), which read in the metadata for each plate from the digitised sample collection forms as well as the results from the hydrolysed and non-hydrolysed assays of each plate. The script then calculated an estimated ODAP content (using the formula described in section 2.2.7, but using a constant factor instead of the slope of a standard curve and without normalising for sample weight) for each sample and generated a single list with the results, containing one line for each sample. Every sample with an estimated ODAP content lower than half the median of its plate was then scored as 'selected', while all other samples were scored as 'rejected'. These results were saved in a column containing '1' for each selected sample and a '0' for each rejected sample.

To narrow down the number of selected samples by removing false positives caused by maternal effects, low sample weight or other factors, all plants that were selected in the first pass of the screen were tested again. By this point, the plants were between five and six weeks old and had grown up into a dense mat. Samples were collected from the youngest, not fully expanded leaves at the top of each plant. Samples that passed the filter again were tested a third time. In the third test, each sample was weighed prior to freeze-drying in order to normalise the reading according to the sample weight.

The high-throughput spectrophotometric assay was prone to a number of artefacts. In particular, specks of dust or scratches on the plates used for the spectrophotometric measurements caused very high spurious readings for some samples. This caused false negatives (high readings in low-toxin samples) when it occurred on the hydrolysed plates and false positives (low readings in normal-toxin samples) when it occurred on non-hydrolysed plates. These false positives presented as individual samples with very low or even negative estimated ODAP contents, while the rest of the family was usually classified as 'rejected'. The list of results was screened manually to remove these events, by inspecting the plates of each sample classed as selected to check whether a sample was present in the well, whether any specks of dust or debris might have interfered with the reading and whether the visible colour of the solution in each well was consistent with the reading provided by the plate reader. Samples that could be ruled out for any of these reasons and individual samples that gave only a borderline positive result with no other samples in the family being classed as 'selected', were curated out.

3.2.7   Confirmation of low-ODAP mutants by testing individual seeds

To measure the β-L-ODAP concentration in seeds from low-toxin mutants identified in the mutant screen, as well as in the progeny of crosses performed to assess gene complementation and to enable pathway analysis, ODAP concentrations in individual seeds were assayed. The same method as described in section 2.2.4 was used. Briefly, seeds were clamped with a peg with rubber pads, orienting the embryo towards the peg. The cotyledons of the seed were carefully drilled using a benchtop drill (Xenox, Föhren, Germany), without breaking the seed or damaging the embryonic axis of the seed. Seed meal was collected into microcentrifuge tubes and freeze-dried. The weights of freeze-dried samples were recorded. For measurements using the spectrophotometric method, 60 % v/v ethanol was used as the extraction medium. ODAP concentrations were measured using the assay method from the original mutant screen. As negative controls, seeds of the *Pisum sativum* variety Cameor (provided by Anne Edwards, JIC) and *Cicer arietinum* flour (bought from Sainsbury's Supermarket in Norwich – variety unknown) were included.

3.2.8   Synthesis of ^{13}C-labelled β-L-ODAP

L-DAP.HCL (50 mg, Sigma-Aldrich, St. Louis, Missouri, USA) was dissolved in 0.9 ml of deionised water. The solution was adjusted to pH 10 using saturated LiOH solution and maintained at 30°C in a glycerol bath. Separately 0.5 ml of ^{13}C diethyl oxalate (Sigma-Aldrich, St. Louis, Missouri, USA) was mixed with 0.54 ml of ethanol. This mixture was added dropwise to the L-DAP solution over the course of two hours, while stirring at 30 °C. The experimental arrangement is shown in Figure 32. Every 10 minutes, the pH of the reaction solution was measured and adjusted back to pH 10 using LiOH solution.

Figure 32. Experimental arrangement for the transeste-
rification reaction. The reaction mixture is stirred using a
magnetic stir bar in the flask. The temperature of the
water bath is automatically maintained at 30 °C.

During this period, the transesterification reaction took place, producing the intermediate ^{13}C-labelled β-ethyl-oxalyl diaminopropionic acid and other products as shown in Figure 33. The reaction was left for another 30 minutes before evaporating to dryness under a vacuum at 40°C in a rotary evaporator (Rotavapor R-215, Büchi Labortechnik, Flawil, Switzerland).
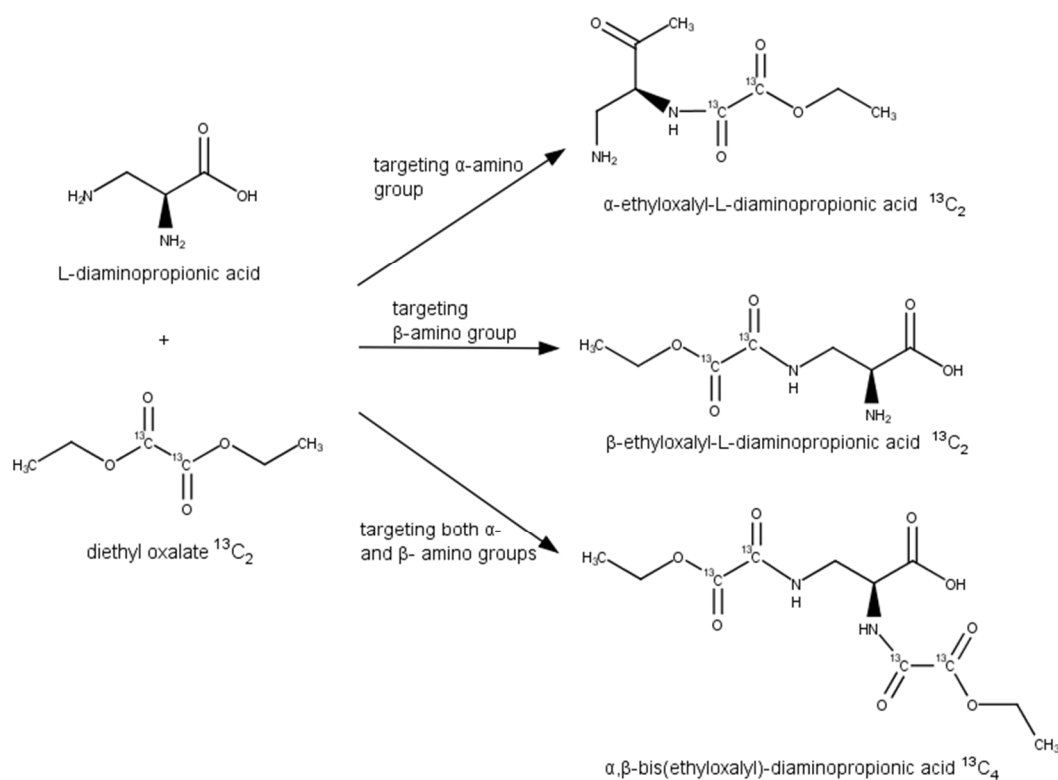
Figure 33. Transesterification reaction forming ^{13}C-labelled β-ethyloxalyl-L-diaminopropionic acid, along with other products. Ethanol is formed as a side product (not shown)

The dry reaction products were suspended in 25 ml of water that had been adjusted to pH 10 using LiOH solution. This solution was then heated to 70°C and incubated over 20h, while being stirred slowly. This resulted in the deprotection of the remaining ester group and the formation of ^{13}C-labelled β-L-ODAP as shown in Figure 34.
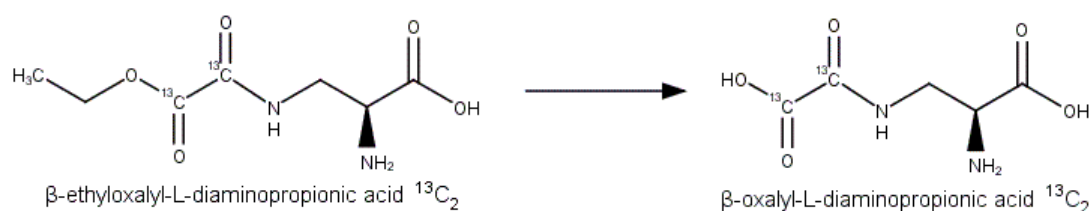
Figure 34. Hydrolysis reaction resulting in the deprotection of the remaining ester group, forming β-L-ODAP. Ethanol is formed as a side product (not shown)

To remove salt, unreacted substrates and unwanted side products from the reaction products, the solution was passed through an ion-exchange column. The column was prepared by washing 9 ml of Dowex 50WX8-400 resin, first with water, then with 1 M HCl

to remove residues of degraded polymer. The column was then flushed with water until the flow-through was back at neutral pH. The reaction products were then passed through the column, followed by 25 ml of water to flush out residues of the reaction products, while retaining salts. The flow through was collected in a 100 ml flask as shown in Figure 35A. The product was freeze-dried (SP Scientific Sentry 2.0, Virtis, Gardiner, New York, USA) overnight and weighed before re-dissolving in 50 ml of water.
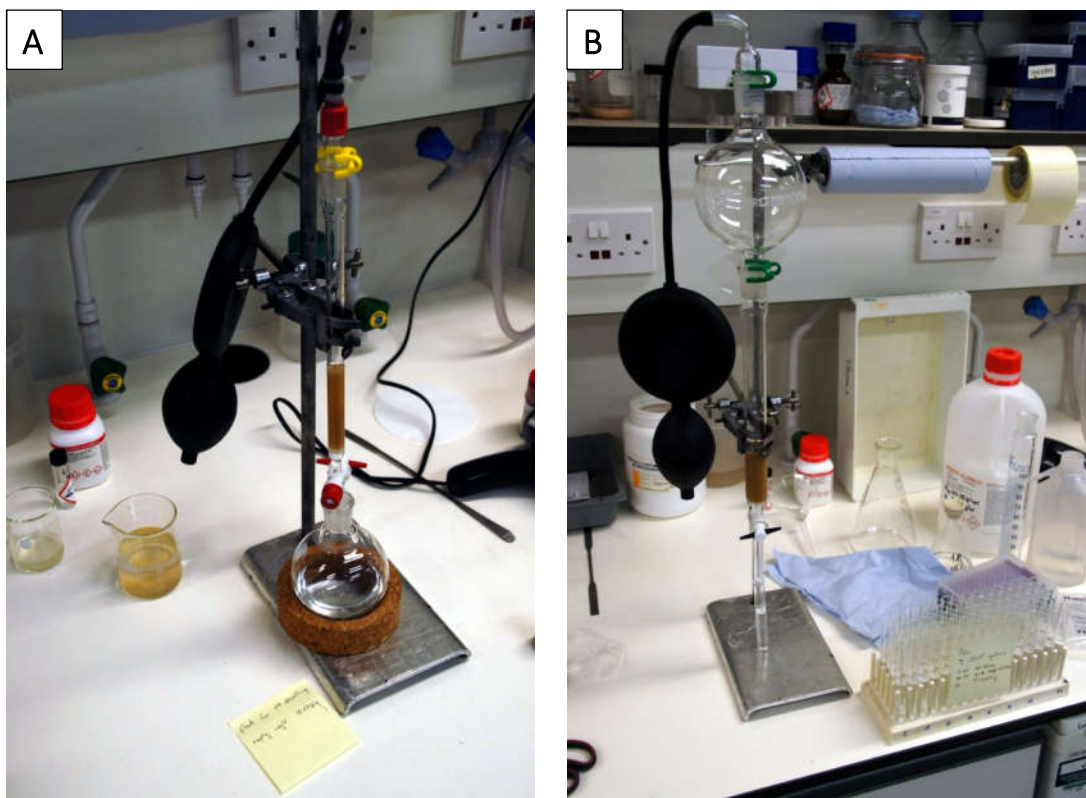
Figure 35. Experimental arrangements of A) the purification and desalting of reaction products and B) their purification and separation into 5 ml fractions.

To remove remaining salts and to separate ^{13}C-β-L-ODAP from the side products of the reaction, the products were passed through an ion exchange column a second time. The column was prepared and washed as before, but using 17.7 ml of Dowex 50WX8-400 resin. The dissolved reaction products (pH < 7) were passed through the column followed by water as shown in Figure 35B. The flow-through was collected into 35 fractions of 5 ml each. The column was then flushed with 300 ml of 0.2 M acetic acid, and 60 fractions of 5 ml each were collected. Samples from each fraction were loaded on a 96-well microtiter plate, and the spectrophotometric assay for ODAP (as described in section 2.2.4) was

performed, measuring the absorbance at 420 nm of both hydrolysed and non-hydrolysed samples to identify which fractions contained L-DAP and ODAP. Appropriate fractions were then combined and freeze-dried. The dry combined fractions were tested using proton and ^{13}C spectra measured using a Bruker Avance III 400 MHz NMR instrument (Bruker, Billerica, Massachusetts, USA) and analysed using the TopSpin version 3.2 software package. The resulting spectra were compared to the spectra of the two isomers of ODAP and 2,3-diaminopropionic acid (Abegaz et al., 1993) to identify the bulked fraction containing ^{13}C-β-L-ODAP.

3.2.9   Characterisation of low-ODAP mutant lines by mass spectrometry

The amounts of β-L-ODAP contained in eight tissues/developmental stages of five genotypes of grass pea were measured using LCMS. The parent variety of the mutant screen, LSWT11, the Indian low-ODAP variety Mahateora and the three mutant lines 1264-2, 4884-2 and 4946-7 were included in this experiment.

Eight tissues were harvested from each genotype:

- leaves of 5-week-old plants grown in a controlled environment room (CER)
- whole roots of 5-week-old plants grown in a CER
- mature, open flowers of 8-week-old plants grown in a glasshouse
- early pods (containing immature seeds) of 9-12-week-old plants, grown in a glasshouse
- late pods (containing immature seeds) of 9-12-week-old plants, grown in a glasshouse
- seeds collected from the plants in the glasshouse
- seedling shoot tips germinating from these seeds (3 days after germination)
- seedling roots germinating from these seeds (3 days after germination)

The classification of pod growth stages (shown in Figure 36) is equivalent to that used by Srivastava and Srivastava (2006).
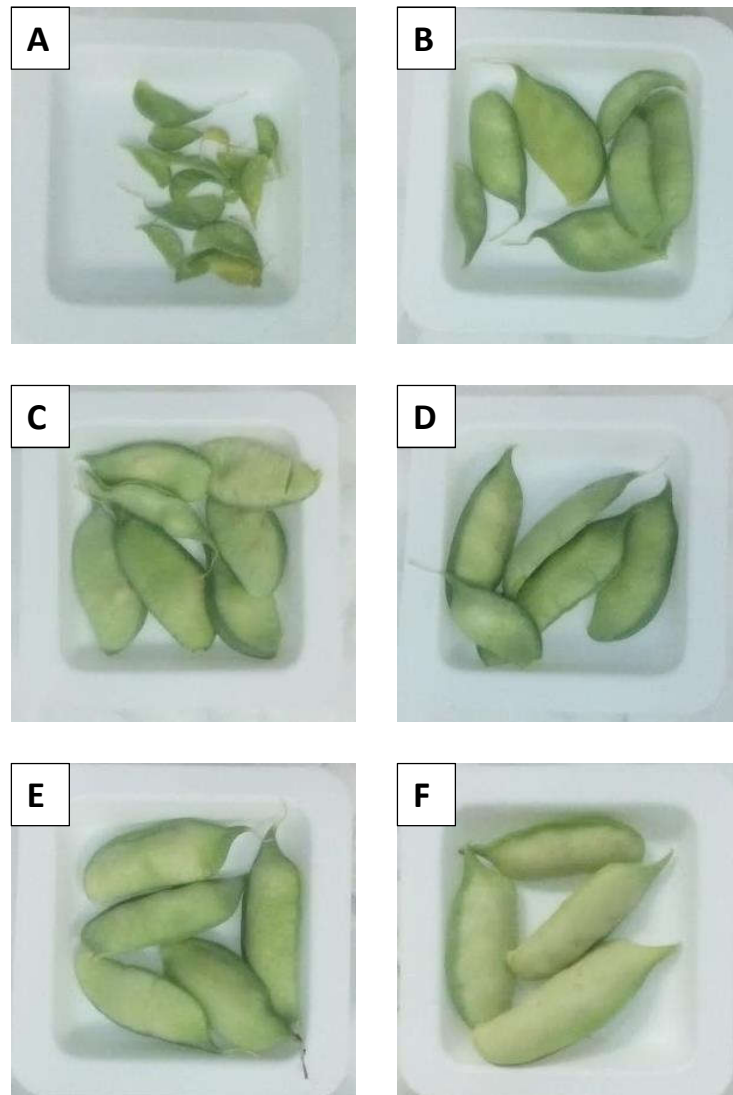
Figure 36. Categorisation of pod growth stages. Equal weights of pods at stages A and B were combined to form the 'early pod' samples. Equal weights of pods at stages C and D were combined to form the 'late pod' samples. Pods at stages E and F were not used in this assay

In the case of seedling shoot and root tips, three samples, each consisting of tissues from between 3 and 6 seedlings were collected. For all other tissues, samples were harvested from six individual plants and bulked. All these tissues, with the exception of seeds, were harvested into liquid nitrogen and then stored at -80 °C. Seeds were harvested when the pods had matured and dried and were stored in dry air at 4 °C.

Leaf, root, flower, early pod and late pod samples were ground using a mortar and pestle, chilled with liquid N_2. To process seed samples, a coffee mill (BISTRO electric coffee grinder, Bodum, Triengen, Switzerland) was used to grind the seeds coarsely before grinding these samples finely using a mortar and pestle chilled with liquid N_2. Care was taken not to let

any samples thaw during grinding. Ground samples were freeze-dried in a BenchTop SLC® freeze drier (Virtis, Gardiner, New York, USA). To process the seedling shoot, and seedling root samples, tissues were freeze-dried before grinding. Samples were ground using one 4 mm steel ball per tube in a TissueLyser ball mill (Retsch, Haan, Germany) shaking at 18 Hz for two minutes. After freeze drying, three replicates of each ground sample were weighed out, aiming for sample weights of 5.5 mg. Accurate sample weights were noted down for later calculation of tissue β-L-ODAP concentrations. Dry samples were stored in sealed tubes at 4 °C.

The ^{13}C-labelled β-L-ODAP standard was added to each sample prior to extraction. The standard was dissolved in deionised water to make a 250 µg/ml solution. An amount of this internal standard was added to each sample, aiming to be within an order of magnitude of the sample's ODAP content. To account for the expected higher β-L-ODAP content in samples of LSWT11, 40 µl of this solution were added to each sample of this genotype, while only 10 µl were added to each sample of Mahateora, 1264-2, 4884-2 and 4946-7. Samples of LSWT11 were diluted 4 times more than the other genotypes after the derivatisation. This allowed the ODAP concentrations of all samples to be measured using the same standard curve.

Extractions were performed using the method described by Kuo et al. for the extraction of free amino acids, including β-L-ODAP, from tissues of Asian ginseng (*Panax ginseng*) (Kuo et al., 2003). To each sample of finely ground, freeze-dried tissue, 500 µl of 70 % v/v HPLC-grade ethanol in RO-water were added and left to extract overnight in a shaking incubator. Samples were centrifuged for 30 minutes at 16,250 g in a Biofuge Pico centrifuge (Heraeus, Hanau, Germany). Supernatants were removed into new tubes and pellets were re-suspended in 500 µl of 70 % v/v ethanol and centrifuged again. Supernatants were added to the supernatants of the first extraction. These steps were repeated a third time. The combined supernatants of all three extraction cycles were evaporated to dryness in a GeneVac (EZ-2 Elite, Genevac, Ipswich, UK). The dried extracts were re-dissolved in 1ml of RO-water.

A series of β-L-ODAP (Lathyrus Technologies, Hyderabad, India) standards was prepared by mixing a 0.05 mg/ml solution and diluting it threefold, seven times. This resulted in a range of standard solutions equivalent to extracts from 5 mg grass pea tissue samples with β-L-ODAP concentrations ranging from 0.00046 % to 1 % of dry weight (w/w). Heavy-isotope-

labelled standard was added to each of these standard solutions to a concentration of 2.5 µg/ml, the same concentration as in the sample extracts.

Sample extracts and β-L-ODAP standards were derivatised using AccQ-Tag™ reagent (Waters, Milford, Massachusetts, USA). The reagent was dissolved in the diluent provided by the manufacturer and used immediately. Derivatisation was performed following the manufacturer's instructions. Briefly, 20 µl of each sample extract and each standard solution were mixed with 60 µl of AccQ-Tag™ borate buffer. To this, 20 µl of dissolved AccQ-Tag™ reagent were added, mixed immediately and incubated at 55 °C for 10 minutes. The derivatised samples were diluted 1:10000 (in three dilution steps) before LCMS injection.

A Xevo triple quadrupole TQ-S instrument (Waters, Milford, Massachusetts, USA) was used to measure the β-L-ODAP contents of samples and standards. To identify suitable mass transitions to create a tuning file, the automatic calibration protocol of the MassLynx software (Waters) was used. For the purposes of calibration, a standard solution containing 10 µM of β-L-ODAP standard, which had been derivatised using the AccQ-Tag™ reagent as described above, was injected directly.

The four major mass transitions, which were identified by this algorithm, were used to detect β-L-ODAP in multi-reaction monitoring (MRM) mode. These mass transitions were

$$347.1 \text{ u} \rightarrow 116.1 \text{ u},$$

$$347.1 \text{ u} \rightarrow 145.1 \text{ u},$$

$$347.1 \text{ u} \rightarrow 171.1 \text{ u and}$$

$$347.1 \text{ u} \rightarrow 303.1 \text{ u}.$$

Of these mass transitions, 347.1 u $\rightarrow$ 171.1 u showed the highest intensity. The mass of the released fragment corresponded to the molecular mass of the derivatisation group. This mass transition was used for quantification of β-L-ODAP. The mass transition 349.1 u $\rightarrow$ 171.1 u was used to measure the internal standard di-^{13}C-β-L-ODAP.

To measure β-L-ODAP contents, a volume of 5 µl of each sample was injected into a Kinetex 2.6 µm EVO C18 100 Å 100 x 2.1 mm column with a C18 guard column (Phenomenex, Macclesfield, UK). The solvent profiles for this experiment are shown in Table 1.

Table 1. Solvent profiles used for the injection of derivatised grass pea extracts to measure β-L-ODAP. Transitions between solvent mixtures were made using linear gradients.

Time (min)	Flow rate (ml/min)	5 mM ammonium acetate in H$_2$O (HPLC-grade) in % v/v	Methanol (HPLC-grade) in % v/v
0	0.4	95	5
8	0.4	90	10
10	0.4	40	60
11	0.4	40	60
12	0.4	92	8
17	0.4	92	8

 A subset of the derivatised samples was also assayed for the presence of the synthesis intermediates O-acetylserine (OAS) and L-2,3-diaminopropionic acid (L-DAP). The injection method used for the β-L-ODAP measurements caused L-DAP and OAS to coelute with other compounds during the flushing of the column with the organic solvent. This necessitated the use of a different injection method to measure these compounds. Because no stable isotope labelled (SIL) internal standard was available for these compounds and only a single dilution of each standard compound was run alongside the samples, rather than a series of standard dilutions that could be used to determine a calibration curve, this assay served as a qualitative measurement of the presence of these compounds in the tissues. The solvent profiles used to test for OAS and L-DAP are shown in Table 2.

Table 2. Solvent profiles used for the injection of derivatised grass pea extracts to measure β-L-ODAP, L-DAP and OAS. Transitions between solvent mixtures were made using linear gradients.

Time (min)	Flow rate (ml/min)	Water + 0.1% formic acid (HPLC-grade) in % v/v	Acetonitrile (HPLC-grade) in % v/v
0	0.6	99	1
0.4	0.4	99	1
7.0	0.4	75	25
8.5	0.4	10	90
9	0.4	10	90
9.1	0.4	99	1
12.6	0.4	99	1

Two mass transitions were used to measure the presence of L-DAP.

275.1 u → 171.1 u and

445.1 u → 171.1 u.

The mass/charge ratios of the two ions prior to fragmentation corresponded to the single- and double-derivatised compound respectively. OAS was measured using the mass transition 318.1 u → 171.1 u.

The ODAP concentrations in analysed samples were calculated using the MassLynx software package version 4.1 (Waters, Milford, Massachusetts, USA) based on the calibration curve (see Appendix 1.2) calculated from the measurements of the series of β-L-ODAP standards. The intensity of the internal standard peak was used to normalise the measurements of samples and standard dilutions.

3.2.10   Genetic analysis by crossing

In order to assess gene complementation between separate mutant lines, crosses were performed. Because grass pea flowers are self-fertile, care must be taken to prevent the stigma of a flower from coming into contact with mature pollen of the same flower. To achieve this, the anthers of the female parent flower were removed before they could dehisce and fertilise the flower. The parts of the grass pea flower are shown in Figure 37. To cross grass pea mutant lines and varieties used as controls, pollen was transferred from a mature flower (the male parent) to the stigma of the emasculated flower using the procedure described below.
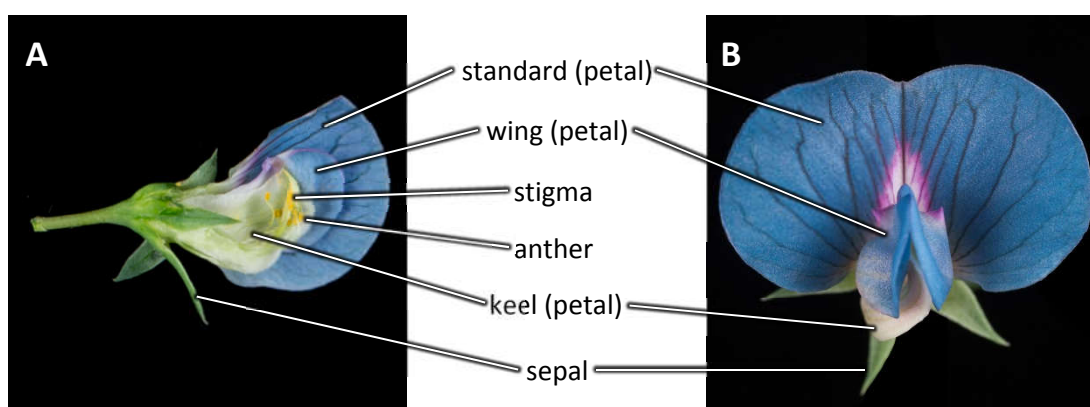
Figure 37. Parts of the grass pea flower. A) side view in cross section B) frontal view of a complete flower

An immature flower, with the standard still joined up at the front and with all petals still showing a yellowish-green colouration was selected as the pollen acceptor (see stages A and B in Figure 38). The recipient flower was marked with a jeweller's tag with the parents of the cross written on it. Using a scalpel, the still-closed standard petal was sliced through from bottom to top to make the inside of the flower accessible. The top and bottom of the flower were gently pressed to open up the wing petals. A small incision was made from the centre to the top of the keel, about 1 mm to the left of the sagittal plane. Using tweezers, the immature pollen sacs and anthers were removed carefully, without damaging the carpel. A mature flower was selected as the pollen donor. Flowers that were used as pollen donors were in stages C and D in Figure 38, just after the standard has opened and the anthers have dehisced. Viable grass pea pollen has a rich golden colour and a powdery texture. Using a scalpel, the pollen donor flower was cut off from the paternal parent plant. The standard and wing petals of the pollen donor flower were removed using tweezers. The base of the keel was grasped with tweezers on the right side, without grasping the enclosed carpel and a tear in the keel was made towards the front of the flower, but not over the pollen sacs. A portion of pollen was removed from the pollen sacs and the hollow at the tip of the keel using tweezers and applied to the stigma at the tip of the pistil of the recipient flower. One pollen donor flower was used to pollinate between 3 and 5 recipients. Before opening a new pollen donor flower, scalpel and tweezers were wiped with 70 % v/v ethanol.
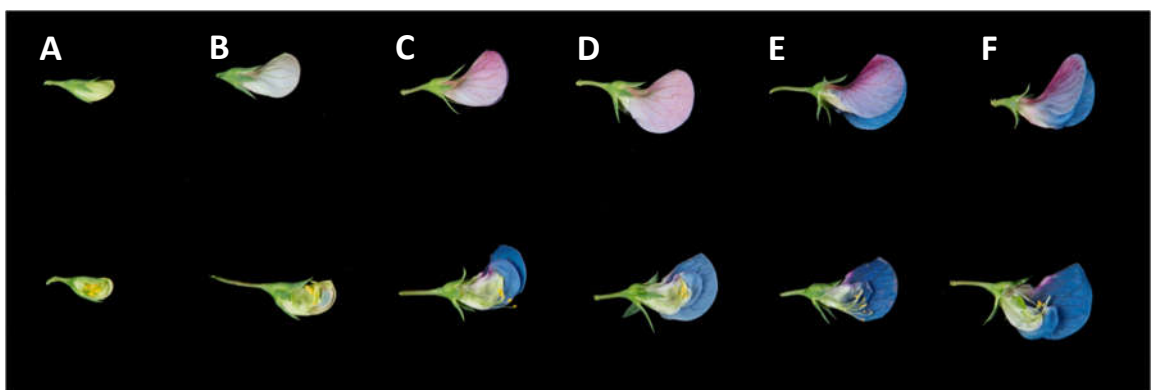
Figure 38. Stages of development of the grass pea flower. The top row shows intact flowers side-on, the bottom row shows flowers with the right half of the petals removed. Flowers remain at each of these stages for roughly one day, but may take several days to fully senesce.

3.3   Results and discussion

3.3.1   Assessment of the mutant population derived from LSWT11

The average 100-seed weight of the M2 families was estimated as 6.05 g (standard deviation 2.32 g) by weighing and counting the seeds in ten seed packets. I used this value to estimate the number of seeds in each packet by weighing and subtracting the weight of the empty paper bag (0.51 g, standard deviation 0.01 g).

The average seed number per family was 141.9 (median = 105.3, SD = 119.1). I selected the 3060 M2 families with the highest number of seeds to be included in the mutant screen. The average number of seeds in each of these families was 199.0 (median = 163.0, SD = 118.4). The distribution of estimated seed numbers per family is shown in Figure 39.
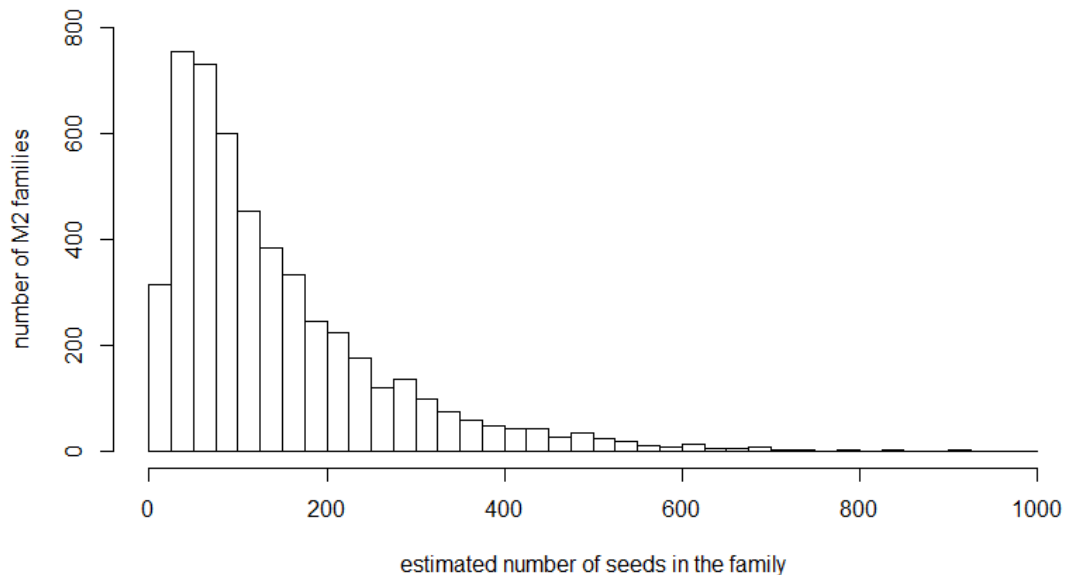
Figure 39. Histogram showing the distribution of seed numbers per M2 family in the mutant population. Seed numbers were estimated based on the weight of each seed packet

Among the 3060 M2 mutant families screened, 26 families (0.85 %) included *albino* and 52 families (1.69 %) included other large-effect chlorophyll mutants, such as *chlorotic*, *chlorina* and *xantha* mutations.

Results on the quality of the mutant population were provided by Ayse Sen and Brad Till of the FAO/IAEA-lab in Seibersdorf Austria (personal communication). The mutation density in the population was estimated to be in the order of one mutation per 850 kbp, based on

four mutations being found in a sample region of 4835 bp among 704 samples. This mutation density is low for a diploid population mutagenised by EMS. It is possible that this result underestimates the true mutation density present in the population, as the analysis was based on two gene targets only, but multiple primer pairs, which may have introduced bias. On the other hand, Sen and Till noted that two of the identified mutations represented the same nucleotide change (C1032Y in the gene PhyA), which is highly unlikely to occur independently in two mutants, and thus might represent heterozygosity. If this was the case it would mean that the true mutation density of the population was even lower. These results correspond to the low number of mutant families containing chlorophyll mutants. The mutation density that can be achieved without reducing the viability of the population below 50 % (a common target value) differs between and even within species (Henikoff et al., 2004). However, the obtained mutation density was very low compared with the densities of most previously described diploid TILLING populations with densities of 1/380 kbp on average (Wang et al., 2012). A population with such a low mutation density would likely reveal very few mutants for every target gene if used for TILLING, unless an unusually large population of DNA extracts was screened every time, increasing the cost of the method. For this reason, and due to the mediocre performance of the parent variety under glasshouse and field conditions in the UK (low germination frequency without scarification as shown below, low yield compared to other varieties – personal communication, Abhimanyu Sarkar, JIC), we decided not to proceed with the production of a TILLING platform based on this present mutant population. A new mutant population, based on the variety LS007 is now being produced for this purpose, but this is not part of the present thesis (personal communication, Abhimanyu Sarkar, JIC).

3.3.2   Optimisation of the mutant screening method

The parent variety LSWT11 showed poor germination frequency without treatment. Seven days after sowing, only 25 % of seeds had germinated (Table 3). Germination frequency was greatly improved by individually chipping the seeds with scissors or scraping them with sandpaper. Both of these methods were equally efficient in promoting germination, but scraping with sandpaper was much faster. Soaking the untreated or chipped seeds in water overnight before planting did not further improve germination frequency.

Table 3. Germination results of LSWT11 seeds subjected to various pre-sowing treatments

Treatment	Seeds sown	Seeds germinated after 7 days	Germination frequency in %
Dry, untreated	12	3	25
Dry, scraped	12	9	75
Wet, untreated	24 (6 imbibed)	8	33
Dry, chipped	24	18	75
Wet, chipped	24 (all imbibed)	17	71

Based on these results, I decided to use sandpaper scraping to scarify the seeds for the mutant screen. To streamline this process, I developed the scarification method using foam mats and pipette boxes lined with sandpaper described in section 3.2.3. This scarification method proved to be significantly better than scratching individual seeds by hand. Combined with the optimisation of the growth conditions, it resulted in a high germination frequency of 96.6 % across the entire mutant screen.

In a preliminary experiment to test the plate-based format for the mutant screen, I assayed a set of 49 LSWT11 shoot samples using the method described in section 3.2.5. I ground up freeze-dried samples and extracted free amino acids by incubating them in 1 ml of water at 95 °C for 30 minutes. I processed aliquots of these extracts with and without the hydrolysis treatment and added the OPA/tetraborate reagent buffer. In addition, I measured a set of mock extractions, which did not include any grass pea tissue. These mock extractions were also subjected to hydrolysis and non-hydrolysis treatments and OPA/tetraborate buffer was added. For the non-hydrolysed mock extractions, water was added instead of 3M KOH solution. Figure 40 shows the results of this experiment. Considerable variation was observed among the absorbance values of hydrolysed samples (mean 0.168, standard deviation 0.033). This variation may be due to genetic variation within the variety, phenotypic plasticity, slight variation in developmental stages assayed, variation in sample size (sample weights were not recorded) and assay noise. Despite this variation, the distribution of absorbance values obtained from hydrolysed samples was clearly separated from the distribution of absorbance values measured for non-hydrolysed samples (mean 0.058, standard deviation 0.004). This experiment showed that the assay using seedling

shoot tip samples produced low background measurements compared to the signal. An ODAP-free sample would result in similar absorbance values after both hydrolysis and non-hydrolysis treatments, making it easy to select from the screening results. Selecting samples with less than half the median ODAP-content as measured by the assay could capture most mutants with strongly reduced ODAP biosynthesis. Samples that were erroneously selected, e.g. due to their low sample size, could be excluded by manual curation and retesting.
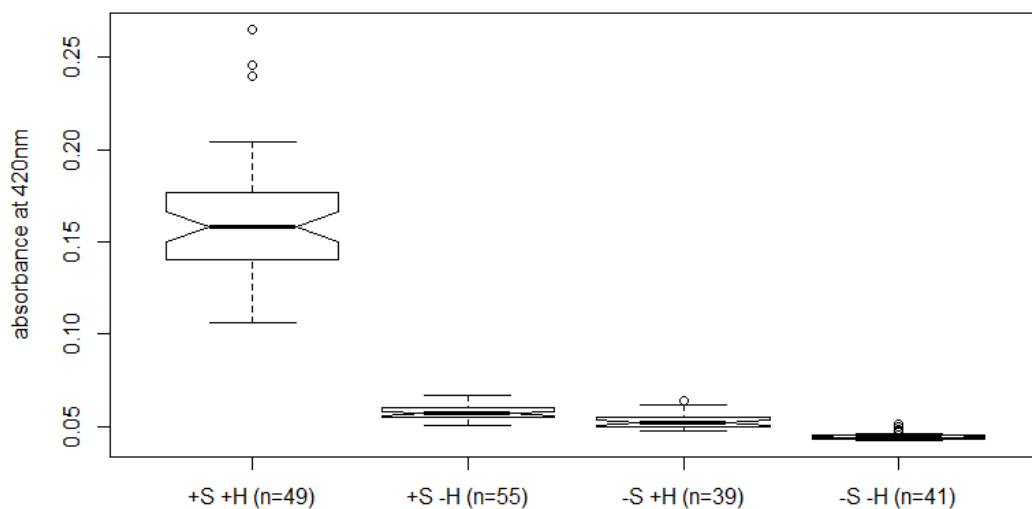
Figure 40. Boxplot showing the absorbance values produced by the spectrophotometric assay run on a set of LSWT11 seedling samples (+S) or mock extractions (-S), which underwent hydrolysis (+H) and non-hydrolysis (-H) treatments. Boxes range from the 25th to the 75th percentile in each distribution, outliers are shown as open circles.

The small, but significant (p<0.0001) difference between hydrolysed and non-hydrolysed mock extractions showed that the presence of KOH solution affects the absorbance reading even without hydrolysis taking place. For this reason I decided to add 3 M KOH to both hydrolysed and non-hydrolysed sample extracts, to prevent absorbance caused by KOH itself being counted as absorbance caused by the reaction of hydrolysed ODAP with OPA and β-mercaptoethanol. To test if the addition of KOH to the sample extracts during the non-hydrolysis treatment caused ODAP to be partially hydrolysed at room temperature before the addition of OPA/tetraborate reagent due to the time needed to pipette reagents on the entire plate, I conducted an additional experiment. Four reaction mixtures were prepared, each in three replicates. Reaction 1 contained 10 µl 0.2 mM β-L-ODAP standard solution, 20 µl dH₂O and 220 µl OPA/tetraborate reagent. Reaction 2 contained 10 µl 0.2 mM β-L-ODAP standard solution and 20 µl 3M KOH solution, which were mixed and left to incubate at room temperature for 60 minutes before addition of 220 µl

OPA/tetraborate reagent. Reaction 3 contained the same constituents as reaction 2, but β-L-ODAP and KOH solutions were left to incubate for only 1 min. Reaction 4 again contained the same constituents but KOH-solution was added after the OPA/tetraborate buffer. After the addition of the OPA/tetraborate buffer, all reactions were left to incubate for 30 minutes at room temperature. Absorbance at 420 nm was measured using a plate spectrophotometer (VersaMax, Molecular Devices, Wokingham, UK). Results of this experiment are shown in Table 4.

Table 4. Absorbance values produced by four variations of the non-hydrolysis treatment.

	Description	**absorbance at 420 nm ± standard error (n=3)**
Reaction 1	No KOH	0.0405 ± 0.0012
Reaction 2	60 min pre-incubation at RT	0.0453 ± 0.0023
Reaction 3	1 min pre-incubation at RT	0.0424± 0.0014
Reaction 4	KOH added after OPA reagent	0.0422 ± 0.0010

As observed before, the presence of KOH in the reaction mixture affected the absorbance values obtained (reaction 4 vs. reaction 1). Leaving β-L-ODAP to pre-incubate with KOH solution at room temperature for 60 minutes resulted in slightly increased absorbance values due to partial hydrolysis, but no notable difference was observed after 1 minute of pre-incubation. Based on these results I decided to add 3 M KOH to sample extracts for non-hydrolysis treatments, followed by addition of the OPA/tetraborate within 1 minute. This sequence of steps (adding the smaller volume first) helped to prevent wells being accidentally skipped during pipetting when performing the highly parallelised assay.

3.3.3   Selection of mutants relative to their plate

A number of non-genetic factors may introduce variation in the spectrophotometric assay that could confuse the mutant screen and lead to type I and type II errors. Among these, some are likely to affect an entire plate of samples that are processed in parallel, rather than individual samples. These include:

-   Harvesting performed by different people, resulting in different average amounts of tissue
-   Harvesting on different days and times of day (risk of physiological effects)
-   Batch differences in buffer mixtures

- Age of OPA-buffer (made fresh every 3-4 days)
- Slight differences in timing of steps during the assay (small effect since the assay uses only end-point reactions)
- Different freezing treatment (time on dry ice, -20°C storage and +4°C after freeze drying)
- Different length of freeze drying (between 23h and 3 days)

To exclude factors that are more likely to cause between-plate variation, rather than between-sample variation, I chose to define the thresholds for the identification of low toxin samples in relation to each 96-well plate, rather than using a single arbitrary threshold for the entire population. In the first pass of the screen, each sample that gave a ODAP value of less than half the median of their plate was selected for re-testing. The same selection method was repeated for the second pass. In the third pass of the screen, samples were weighed first in order to normalise the results by tissue amount. Only plants that gave an average reading of less than two standard deviations below the control average were selected.

To exclude false positives, i.e. plants that were wrongly identified as low-toxin, all plants that are selected as low-toxin in the first pass of the screen were sampled again (taking young leaf tissue from juvenile plants).

3.3.4   Scarification and priming

In preliminary experiments, LSWT11 seeds and seeds of the mutant population germinated poorly without prior treatment. This could have presented a problem for the mutant screen as seeds that failed to germinate would have resulted in missing data, necessitating re-screening of many mutant families. Screening twelve seeds of an M2 family should give a nearly 97 % chance of including at least one homozygous mutant, provided the homozygous mutant allele does not cause seed abortion:

$$\left(1 - \left(\frac{3}{4} \right)^{12} \right) \, 100 \, \% \; = 96.8 \, \%$$

I therefore devised a method for rapidly scarifying all the seeds from one mutant family together, in order to increase the frequency of germination substantially. Scarifying boxes and blocks lined with sandpaper, as used to scarify Medicago seeds (Garcia et al., 2006),

cannot be used easily to scarify grass pea seeds, because of the compact wedge shape of these seeds (especially for small-seeded varieties). Due to this shape, seeds could slip out or escape scarification if they were slightly smaller than other seeds being scarified at the same time. To get around this problem, I developed a scarifying method where seeds of different sizes could be held in place at the same time.

3.3.5   Data handling and selection of putative mutants

In order to process the large amount of data created by the mutant screen I relied on computational processing to identify putative mutants. For this purpose I wrote a HTML data entry form to digitise the metadata generated during sample collection and a script using the statistical programming language R, which processed metadata and spectrophotometric measurements to select putative mutants from the dataset. I manually curated these results to remove false positives and selected a subset of these individuals for repeated tests using the spectrophotometric method. In total, the mutant population underwent three rounds of testing to narrow down the number of putative mutants.

3.3.6   Low-toxin mutants identified

Both pilot and main screens revealed low-toxin mutants that I investigated further by re-testing to exclude false positives[3]. Many of the mutant families with low-toxin mutants contained more than one individual with abnormally low toxin content. The numbers of putative mutants identified in the pilot and main screens and the numbers of mutants that were confirmed in the second and third pass of the mutant screen are shown in Table 5.

---

[3] A substantial part of the experimental work for this mutant screen has been performed by Kalyani Kallam and Hsi-Hua Wang (both at JIC), under my supervision. All analysis is presented here was performed by me. These data have not been submitted elsewhere

Table 5. Summary of results of the mutant screen, using the high-throughput spectrophotometric method. In each cell the total number of screened M2 families or individuals is shown. Numbers in brackets are the numbers screened in the pilot and main screens, respectively. * samples were not weighed before the assay ** samples were weighed before the assay to normalise the measurement by the amount of sample taken, giving a more accurate result for toxin levels

	First pass screen	**Second pass screen**	**Third pass screen**
Type of screen	Not normalised*	Not normalised*	Norm. by weight**
Number of M2 families screened	3060 (770 + 2290)	314 (90+224)	52 (19 + 33)
Number of plants screened	36696 (9240 + 27456)	625 (176 + 449)	97 (29 + 68)
Number of M2 families that passed the filter	314 (90 + 224) (after manual curation)	52 (19 + 33)	37 (4+33)
Number of M2 plants that passed the filter	625 (176 + 449) (after manual curation)	97 (29 + 68)	68 (6 + 62)

Using the high-throughput assay, I successfully identified putative low-ODAP mutants from the mutagenized population. The limited sensitivity of the spectrophotometric assay meant that I could not infer whether any of these mutants were entirely ODAP-free from these data alone.

3.3.7   Confirmation of low-ODAP mutants

The plants identified as low-ODAP mutants that passed through three rounds of testing during the pilot screen were transferred into individual pots eleven weeks after sowing and left to mature and set seed. The same was attempted during the main screen, but most of these plants did not establish well in the new pots and died without producing any seeds. High temperatures in the glasshouse, insufficiently deep pots and high pressure of thrips and powdery mildew may have contributed to these problems.

In order to identify new putative mutant plants to replace those lost from the main screen, as well as to confirm the originally identified mutant families, I sowed out 12 additional seeds from the original packets of each of the 33 M2 families selected in the main screen, which had been scored as low-ODAP in the three passes of screening. These seeds were scarified and sown into deep-rooting pots. Seedling shoot tips and top leaflets of these

plants were harvested and assayed using the spectrophotometric method to identify low-ODAP mutant plants. Results of these experiments are shown in Figure 41. The correlation of ODAP concentrations in seedling shoot tip and top leaflet samples from the same plants is shown in Figure 42. This also narrowed down the list of mutant families, as not all of them were again found to contain low-toxin mutant plants. Based on these data, individual mutant plants were selected for further experiments.
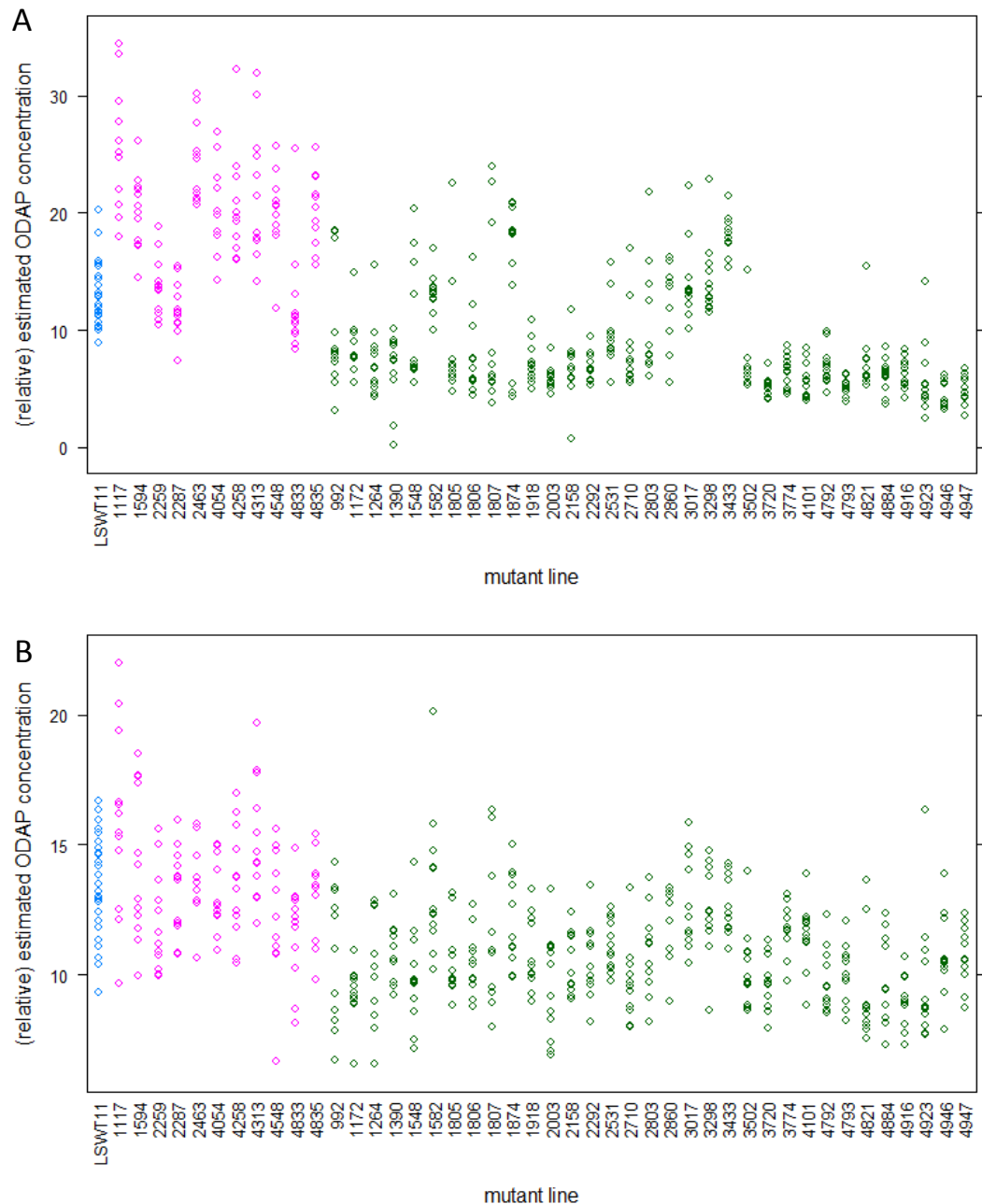
Figure 41. ODAP concentrations of the replacement plants of M2 families selected during the main screen. No standard series was included in these experiments, hence ODAP concentration in the shoot tip samples could not be calculated in absolute numbers. The estimated ODAP concentration is therefore given in arbitrary units. A) seedling shoot samples, B) top leaflet samples of six-week-old plants
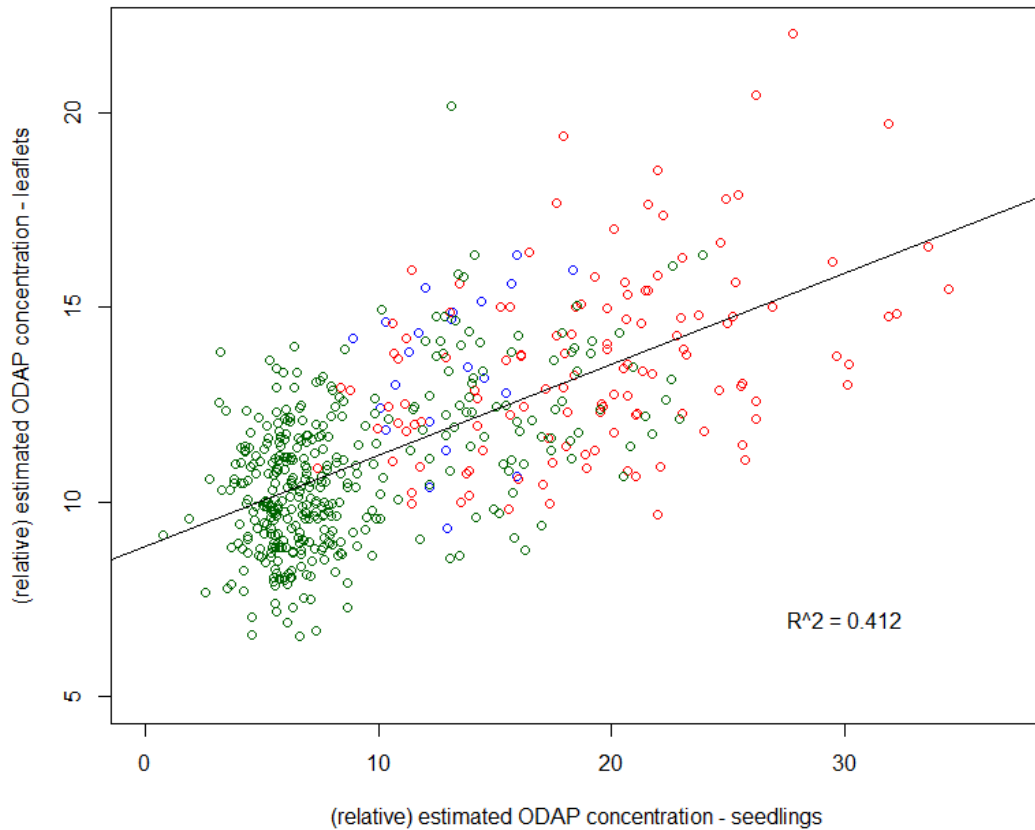
Figure 42: Correlation of ODAP concentrations in seedling shoot tips and leaflets of mature plants

Individuals with very low-ODAP concentrations in either the seed or seedling screens and low-ODAP outliers compared to the rest of their M2 families were selected for further investigation. Overall, individuals from 13 M2 families from the main screen, and individuals from 4 M2 families from the pilot screen (not included in Figure 41) were chosen to form individual mutant lines. The name of each mutant line was composed of the ID of the M2 family and the ID of the individual low-ODAP mutant plant that was selected out of the twelve individuals screened in the family, e.g. the offspring of plant two of the M2-family 490 were designated as mutant line 490-2. Seeds produced by the selected plants were scarified and sown. Samples of seedling shoot tip tissue were collected and analysed for their ODAP concentrations using the spectrophotometric method. Results are shown in Figure 43. The mutant lines 627-7, 3298-4, 1264-1, 992-4 and 1874-5 did not reproduce the low-ODAP phenotype in this generation and were discarded.
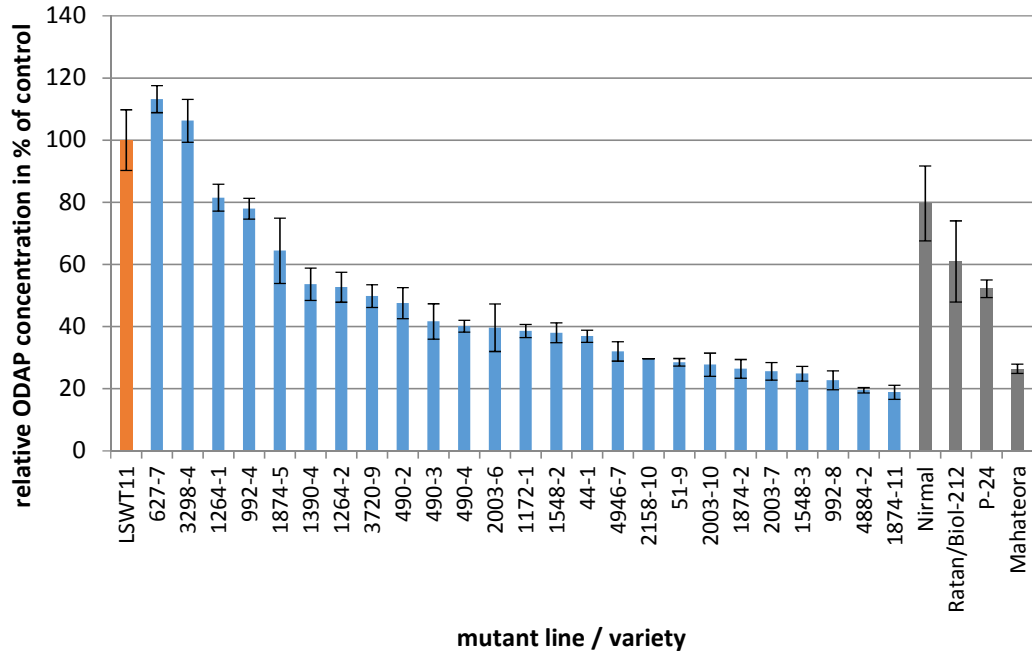
Figure 43. ODAP concentration in young shoots of selected mutants and controls and varieties, measured by the spectrophotometric method. Results are displayed in percent relative to the parent variety LSWT11. Error bars denote standard error based on three biological replicates.

The lines 44-1; 51-9; 490-2,3 and 4; 992-8; 1172-1; 1264-2; 1390-4; 1548-3; 1874-2 and 11; 2003-6 and 7, 2158-10, 3720-9, 4884-2 and 4946-7 were retained and used for crossing experiments. Some flowers of these plants were allowed to self-fertilise to produce seed. Seed meal from individual seeds was assayed for ODAP concentration using the spectrophotometric method. Results are shown in Figure 44.
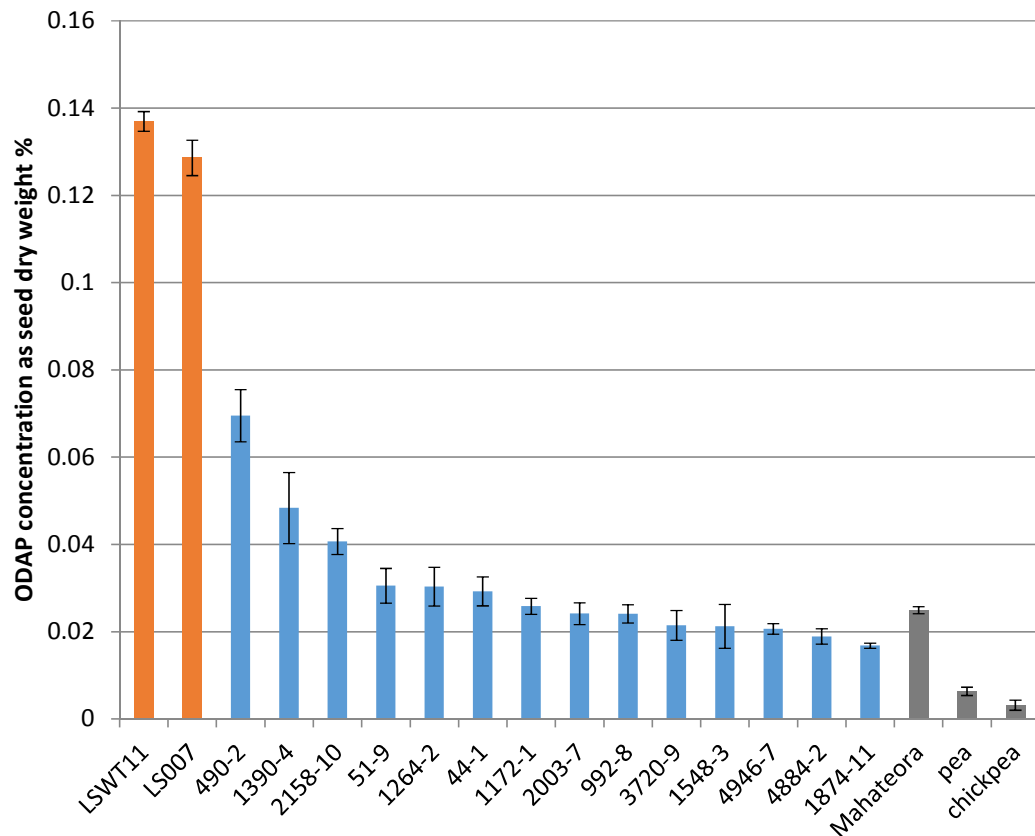
Figure 44. Seed ODAP concentrations of varieties and mutant lines, as well as pea and chickpea seed meal samples, measured by the spectrophotometric method. Error bars denote standard error based on five seeds analysed.

The selected mutant lines showed strongly reduced ODAP concentrations in seeds, seedling shoots and leaf tissue (second and third pass of the mutant screen) compared to the parent variety, which was reproduced in the following generation (shoot tip results shown in Figure 43, seed results shown in Figure 44). The lowest of these values were in a similar range to previously released low-toxin varieties, such as Mahateora (Bhowmick, 2013), Ratan (Santha and Mehta, 2001) or Ceora (Siddique et al., 2006). It is conceivable that the newly described mutations could be introgressed into existing low-ODAP varieties to further reduce their β-L-ODAP content. The parent variety used in this screen had a high ODAP content, representative of the resilient, small-seeded landraces that are often grown in subsistence agriculture. This variety was selected to make low-ODAP mutants easier to identify against a high-ODAP background. In the genetic background of an already low-toxin elite variety, these mutations might cause a further decrease in toxin content. The mutants thus represent a significant advance in the breadth of genetic material for the breeding of low-toxin grass peas. In addition, they may represent mutations that could help to reduce the environmental sensitivity of toxin production that complicates the breeding

of reliable low-toxin varieties (Fikre et al., 2006; 2011). Further tests involving the growth of the mutant lines under different environments known to increase toxin production, such as high Fe^{3+}, low Zn^{2-}, drought or high salinity, will be needed to ascertain whether any of the mutants represent useful traits in this regard.

However, the mutant screen did not reveal any mutations that led to the complete loss of toxin synthesis. This result could be for one of, or a combination of, several reasons:

- No null allele was generated in the mutagenesis. The mutation density of the population was lower than what was aimed for, as shown by the low frequency of chlorophyll mutations and the data provided by Sen and Till. It is possible that no mutation occurred in the entire population that could render a gene product essential for the synthesis of β-L-ODAP completely non-functional.

- β-L-ODAP is essential for the survival of the plant. If the synthesis of β-L-ODAP or an intermediate were essential for the plant, any mutation that stops the synthesis of β-L-ODAP, or the relevant intermediate, entirely would render the plant inviable (i.e. the mutation is homozygous lethal). In this scenario all mutants that knock out toxin synthesis entirely would be missing from the analysis.

- The genes involved in β-L-ODAP biosynthesis and its regulation are redundant. If more than one gene encoding enzymes for each step of the biosynthetic pathway were present, then no single mutation in a gene encoding a biosynthetic enzyme would knock out synthesis entirely. If loss of any duplicated genes causes less than a 50 % drop in toxin synthesis, this might not have been picked up in the mutant screen, due to the inherent inaccuracy of the screening method. Thus even recombining the identified mutants might fail to reduce the toxin levels to zero, as the remaining activity would be due to small-effect genes that the initial mutant screen did not pick up. Similarly, if the regulation of β-L-ODAP synthesis occurs through redundant regulatory pathways, then even the complete knockout of one of the regulatory pathways that activate production might not reduce toxin production to zero. Instead, it might cause only a reduction of toxin synthesis under specific environmental conditions, potentially leading to unexpectedly high toxin levels in different environments.

- The β-L-ODAP content measured in the screen is swamped by a maternal effect. If β-L-ODAP itself, or an intermediate of its synthesis were produced in the mother and deposited in the developing seed and then translocated into the shoot tip during germination, it may have confused the results of the first stage of the

mutant screen, where very young tissues were used. This could lead to the β-L-ODAP content of seedlings to be overestimated if the seedling is homozygous for a mutation that reduces toxin synthesis, while its mother plant is heterozygous, resulting in low-/zero-ODAP mutants being erroneously rejected.

It is also possible that several of these factors may be at play at the same time. For example, some genes involved in the biosynthesis pathway of β-L-ODAP may be essential, especially the earliest steps and the synthesis of oxalyl-CoA, because they are simultaneously involved in different, essential, cellular processes, while the genes responsible for the later steps of β-L-ODAP synthesis, such as the synthesis of L-DAP and β-L-ODAP itself, might be duplicated.

3.3.8 Synthesis of a heavy-isotope-labelled internal standard for LCMS

To quantify the amount of β-L-ODAP contained in these samples accurately using LCMS, it was necessary to include an internal standard that was chemically similar to β-L-ODAP, but could be distinguished by its mass spectrum. Stable-isotope-labelled (SIL) compounds, which contain atoms of uncommon, but stable isotopes, are well-suited for this as their chemical properties are nearly identical to the compound to be measured, resulting in both compounds undergoing the same chemical reactions (Stokvis et al., 2005). By adding a known quantity of SIL-β-L-ODAP to a sample with unknown β-L-ODAP content prior to extraction, it is possible to normalise the result of an LCMS experiment according to the efficiency of the extraction and derivatisation. With a higher number of atoms in a molecule, the chances of it containing naturally occurring heavy isotopes are increased. Depending on the size of the molecule in question, a greater number of heavy isotope atoms in the molecule are necessary to differentiate the mass spectrum of the labelled compound from the naturally-occurring heavier molecules reliably. Based on the sum formula of β-L-ODAP ($C_5H_8N_2O_5$) and the relative abundances of heavy isotopes of carbon, hydrogen, nitrogen and oxygen in the environment, expected ratios of β-L-ODAP isotopic homologs of different molecular weights can be calculated, assuming specific isotopes are not strongly selected for during the biosynthesis of β-L-ODAP. The mass of β-L-ODAP containing only the most common isotopes is 176.043 u. For every molecule of this mass, 0.06 molecules of approximately 177 u and 0.01 molecules of approximately 178 u are likely to exist in grass pea samples due to these naturally occurring heavy isotopes. To

differentiate the internal standard from naturally occurring heavy isotopic homologs, the incorporation of a mass of at least 2 u was necessary.

In order to enable accurate measurement of β-L-ODAP using LCMS, I thus tried to synthesise β-L-ODAP labelled with two carbon-13 atoms as an internal standard. I started by synthesising unlabelled β-L-ODAP to establish a synthesis technique that I could use to synthesise the labelled compound by substituting a ^{13}C-labelled substrate. My first attempt was to synthesise dimethyl oxalate by combining oxalic acid dihydrate with (trimethylsilyl)diazomethane (TMS-diazomethane) in diethyl ether. This was unsuccessful as the reaction products were contaminated with by-products. I tried to resolve this issue by azeotropically removing the crystal water in the oxalic acid dihydrate and adding TMS-diazomethane in excess. This was successful, but the yield of this reaction was too low to continue to the next step of the reaction. I therefore tried a different approach to synthesising dimethyl oxalate (Jaoui et al., 2004). For this reaction, I combined oxalic acid dihydrate with methanol in the presence of $BF_3.Et_2O$ as a catalyst. Despite several steps of separation and extraction many contaminants were left in the reaction products. I abandoned this synthesis approach and instead attempted to reproduce a previously described one-step synthesis method for β-L-ODAP (Harrison et al., 1977), following the method described in section 3.2.8. This method was successful in producing β-L-ODAP.

To produce di-^{13}C-labelled β-L-ODAP, I followed the same process, but using diethyl oxalate [13]C_2 (Sigma-Aldrich, St Louis, Missouri, USA) instead of unlabelled diethyl oxalate. After an initial purification and desalting step using an ion exchange column, I passed the reaction products through a second ion exchange column, flushed the column with water and 0.2 M acetic acid and collected 95 fractions of 5 ml each. I then sampled each fraction and performed the spectrophotometric ODAP assay with and without prior hydrolysis. As shown in the results from hydrolysed samples in Figure 45, products that hydrolysed to L-DAP eluted from the column in three phases. The fractions 16 to 22 also showed slightly higher than background absorbance values when non-hydrolysed samples were used. Based on these data, I pooled the fractions in four groups: F I (fractions 2 to 13); F II (fractions 16 to 22); F III (fractions 25 to 49) and F IV (fractions 50 to 69). The fractions in between (14, 15, 23 and 24) were not included in these pooled fractions to reduce cross-contamination.
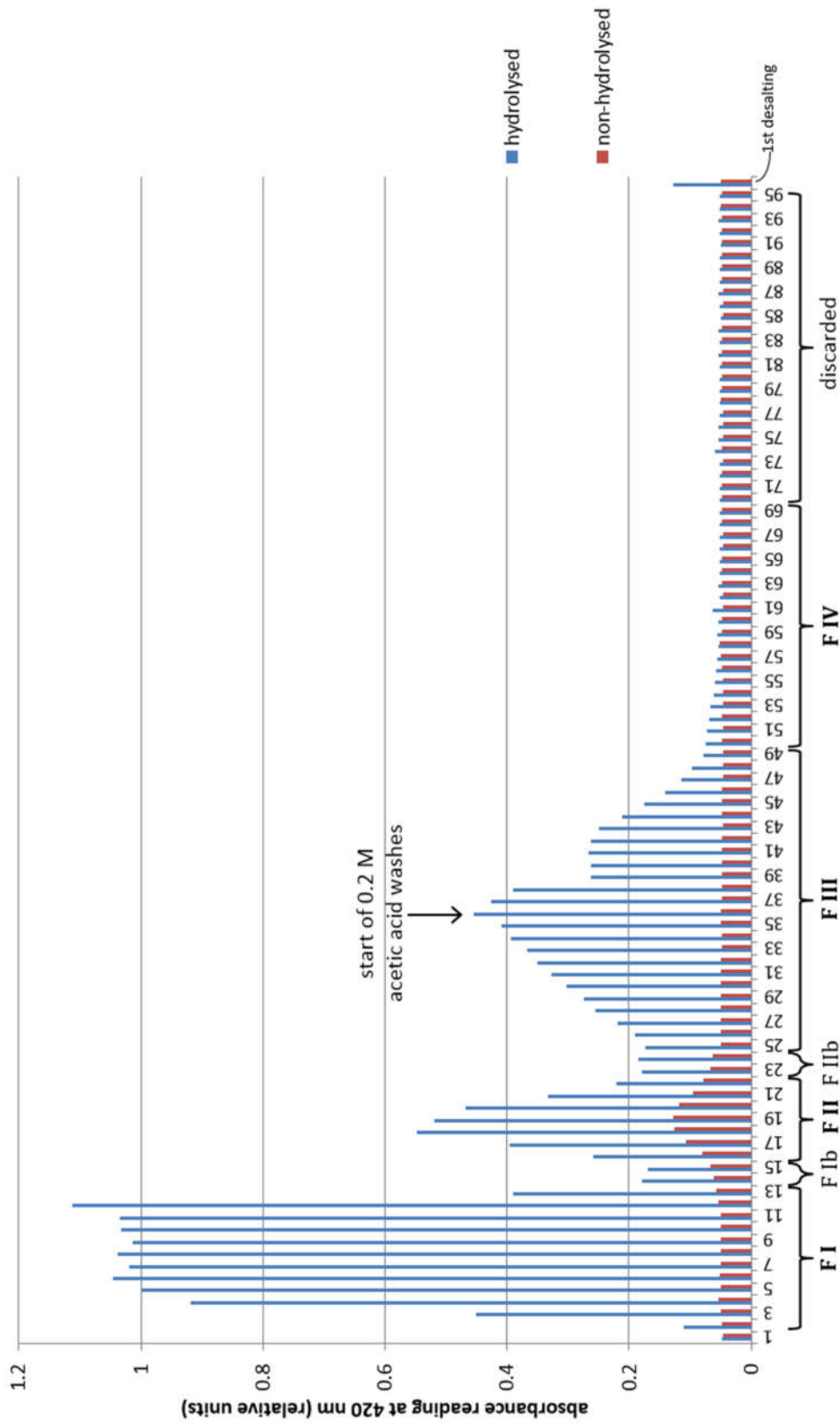
Figure 45. Absorbance measurements produced by the spectrophotometric assay for ODAP in 5 ml fractions obtained by fractionation using a Dowex 50WX8-400 ion-exchange column. Raw measurements (without blanking) are shown for both hydrolysed and non-hydrolysed samples. Based on these results, the fractions were combined into four bulked fractions F I, F II and F III, labelled F Ib and F IIb, respectively, were not included in the fractions to allow a cleaner separation of compounds. The last sample in this series shows the products after the first purification/desalting step, prior to fractionation.

The identity of the reaction products in the pooled fractions was confirmed using NMR spectroscopy. Figure 46 shows the structure of β-L-ODAP for reference. The proton-NMR spectra of the first three pooled fractions and an unlabelled β-L-ODAP standard are shown in Figure 47. Fraction F IV was not included in this experiment because it did not appear to contain significant amounts of product (see Figure 45). Because the samples were dissolved in D_2O, hydrogen atoms in hydroxyl, amino and imino groups are replaced with deuterium atoms, rendering them NMR-inactive. Hence, only the hydrogen atoms attached to the carbon atoms 2 and 3 produced peaks in the proton NMR.
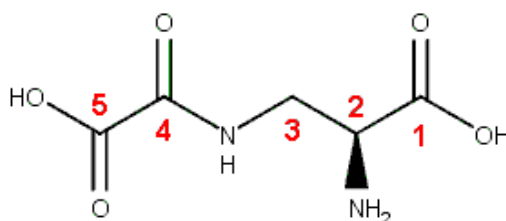
Figure 46. Structural formula of β-L-ODAP. Carbon atoms are numbered in red. In the heavy-isotope labelled isoform, carbon atoms 4 and 5 are replaced with carbon-13 atoms. Due to the chirality of the carbon atom 2, the two hydrogen atoms bound to carbon atom 3 produced separate peak groups in the NMR spectrum (3a and 3b)
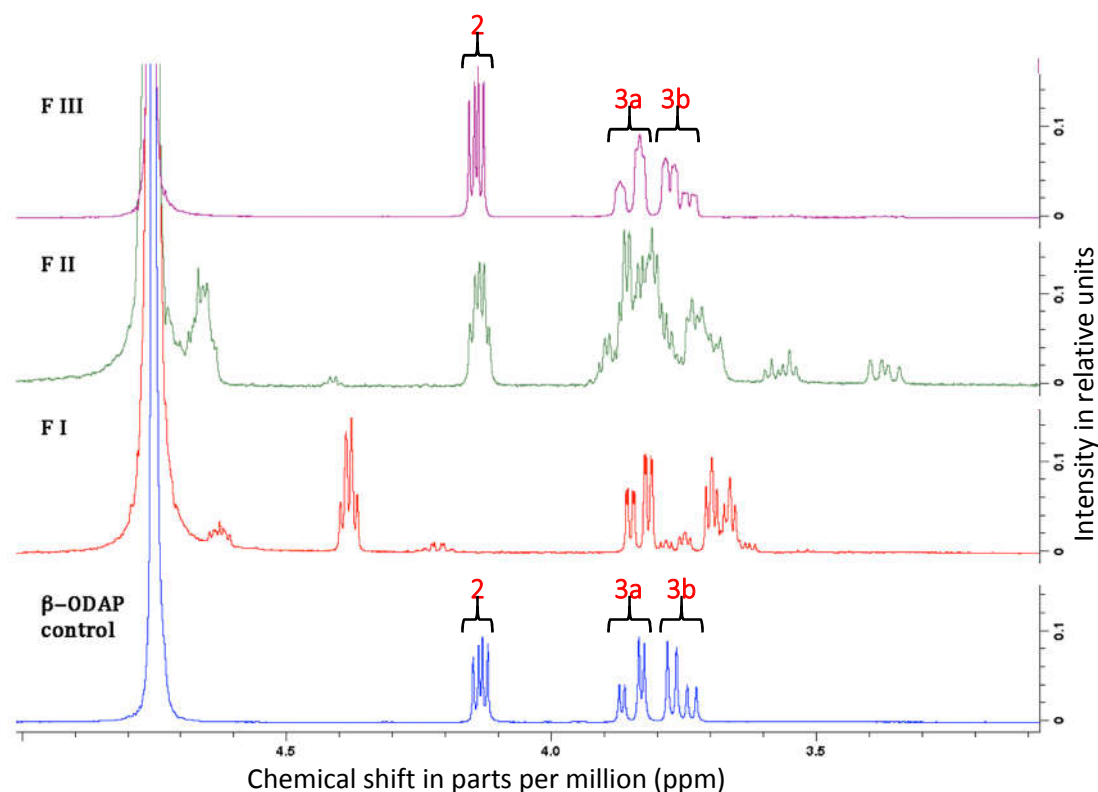
Figure 47. Proton NMR spectra of the pooled fractions F I, F II and FIII of reaction products of the SIL-β-L-ODAP synthesis and unlabelled β-L-ODAP. Samples were dissolved in D₂O and measured using a Bruker 400 MHz instrument.

All spectra contained a dominant HDO-peak (water containing one ^{1}H and one ^{2}H atom) which was used to align the spectra. The spectrum of Fraction F I did not match the spectrum of the β-L-ODAP sample. However, three main peak groups in this spectrum (a doublet of doublets at 4.39 ppm, a doublet of doublets at 3.84 ppm and a doublet of triplets at 3.67 ppm) could indicate that this sample contained α-L-ODAP. Other groups of peaks present in this spectrum were likely due to other by-products of the reaction. The spectrum of Fraction F II contained several complex multiplets, indicating that probably F II contained several compounds. As shown in Figure 45, the individual fractions that were combined to form F II gave higher-than-background readings in the spectrophotometric assay without hydrolysis. This suggested the presence of leftover L-DAP that did not react. Bis-oxalyl-α,β-L-diaminopropionic acid may be another by-product contained in this fraction. The proton-NMR spectrum of fraction F III contained a doublet of doublets centred at 4.13 ppm and two doublets of doublets centred at 3.75 ppm and 3.85 ppm respectively. This corresponded to the spectrum of the β-L-ODAP control and to the proton-NMR spectrum of β-L-ODAP recorded in the literature (Abegaz et al., 1993). The spectrum of Fraction F III also contained a singlet peak at 2.07 ppm (not shown), which was

most likely due to contamination with a volatile organic solvent (methanol or acetone) left in the NMR-tube, rather than a contamination of the dry sample. These data indicated that F III contained the desired product.

To confirm that Fraction F III contained ^{13}C-labelled β-L-ODAP, a carbon-13-NMR was run on the F III and β-L-ODAP standard samples. The results are shown in Figure 48. The chemical shifts of the peaks corresponding to carbon atoms 1, 2 and 3 were identical in both spectra. The peak corresponding to two carbonyl groups in positions 4 and 5 was barely discernible against the background noise in the β-L-ODAP standard, but was very prominent in the spectrum of F III. This indicated an extremely high rate of ^{13}C-atoms in these positions.
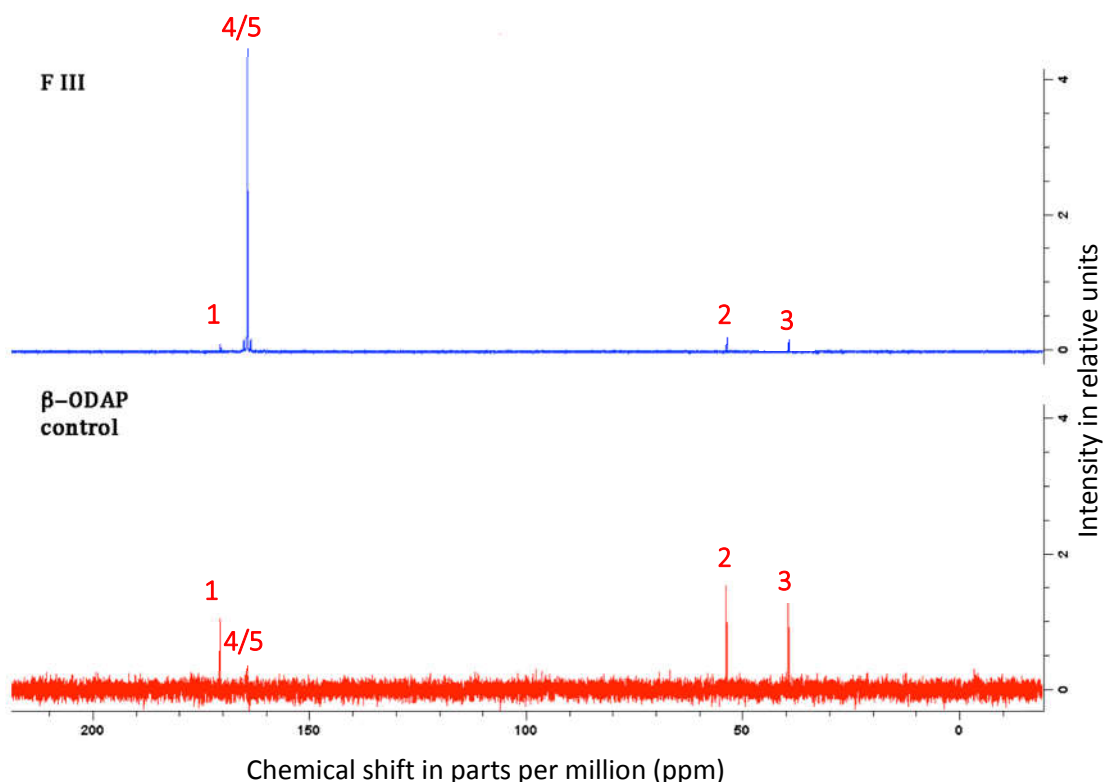
Figure 48. Carbon-13-NMR spectrum of Fraction F III and β-L-ODAP standard. Carbon atoms are labelled as in Figure 46. The spectrum of F III is scaled down to accomodate the large peak resulting from the ^{13}C atoms in positions 4 and 5.

These results showed that Fraction F III contained di-^{13}C-labelled β-L-ODAP, with no visible contaminants. After freeze-drying this fraction, 4.2 mg of residue were retained. This was only 6.7 % w/w of the theoretical yield of 62.6 mg, if all of the reaction substrates had been

turned into di-^{13}C-β-L-ODAP. However, the obtained amount was of high purity and sufficient for my subsequent LCMS experiments.

3.3.9   Confirmation and characterisation of mutants

Having identified putative mutants with low ODAP contents using the spectrophotometric screen, I needed to confirm these data using an independent method of measurement. Methods using High Performance Liquid Chromatography (HPLC) have previously been described to measure β-L-ODAP contents (Zhu et al., 2006; Kuo et al., 2003; Fikre et al., 2008; Ghosh et al., 2015). Separation of extracted free amino acids on a HPLC column allows fluorometric measurement of ODAP. These methods were not applicable for use in a mutant screen, because they were not as easily scalable as the spectrophotometric method. However, they allowed for more accurate measurements. HPLC methods are also able to separate the two isomers of ODAP to measure the ratio between α- and β-L-ODAP in a sample. The accuracy and sensitivity of this method can be further improved by coupling the chromatographic separation of compounds with mass analysis using mass spectrometry. Because seeds are the most commonly-used tissue for ODAP analyses and the most important tissue for human consumption, measurements of ODAP concentration in seeds are ultimately more relevant than the ODAP concentration in young shoot tips or leaves. I therefore screened meal samples from seeds produced by putative mutant plants using an LCMS method developed for this purpose. This was especially useful in handling samples from low-ODAP mutants and varieties as it allowed for very low limits of detection.

To ascertain whether the low-toxin phenotype apparent in the young shoots of the mutant lines also affected the toxin concentration in the seeds, I tested seed meal extracted from individual mutant seeds by drilling into individual seeds as described in section 2.2.5 and measuring the β-L-ODAP concentration using LCMS. In this preliminary experiment, the internal standard described in the previous section was not used, and instead I relied on a series of external standards (see calibration curve in Appendix 1.2). Apart from the use of external standards only, the methods described in section 3.2.9 (solvent profile shown in Table 1) were used. The results of this experiment are shown in Figure 49. These data confirmed the relative levels of ODAP in LSWT11 and the mutant and low-ODAP lines, but LCMS measurements of ODAP were overall much higher than the previous spectrophotometric measurements in individual seeds described in section 3.3.7. However, the averages produced by the varieties LSWT11, P-24, Mahateora and Nirmal were similar to the β-L-ODAP concentrations measured in bulk seed samples by the spectrophotometric

method as shown in section 2.3.4 (Ratan was not measured in the bulk seed meal experiment).
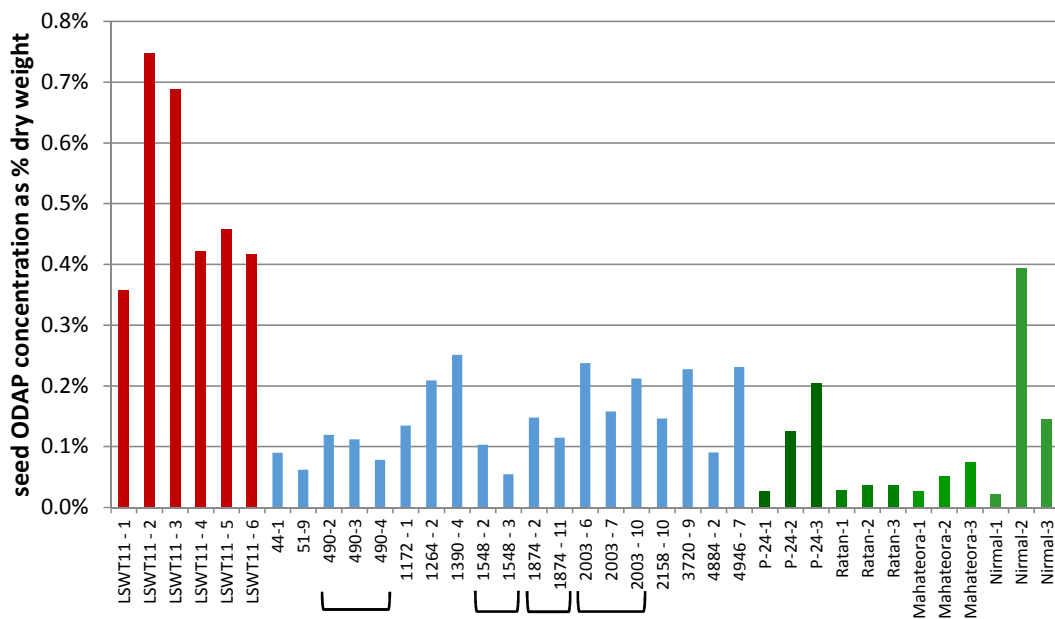
Figure 49. ODAP concentrations in individual M3 seeds of selected mutant lines and controls, measured by mass spectrometry using a series of external standards. Results from individual seeds are shown to display the high seed-to-seed variation observed in the control varieties. Shown in red is the parent variety LSWT11, blue the selected low-toxin mutants and green the four Indian low-toxin cultivars. Brackets denote samples of M3 seeds from the same M2 families.

Large variations were seen between the ODAP concentrations of individual seeds, especially among the parent variety LSWT11 and the low-ODAP varieties P-24 and Nirmal. Similarly large variations between individual seeds of the varieties were seen using spectrophotometric measurements, as described in section 2.3.2. The variety Ratan, which gave highly variable results in the experiment described earlier, gave consistent results among the three seeds included in this experiment. The variety Mahateora showed low ODAP concentrations with little variation between seeds in both this experiment and the spectrophotometric measurements. Four mutant families included seed samples from more than one individual that was selected in the mutant screen: 490, 1548, 1874 and 2003. The high variability between seeds may be due to genetic diversity within varieties or phenotypic plasticity.

I also used LCMS to measure levels of β-L-ODAP in eight tissues of five genotypes of grass pea (the parent variety of the mutant screen, the low-ODAP variety Mahateora and three putative mutant lines identified in the screen), allowing me to further characterise these mutant lines. To reduce the variability associated with individual seed samples, I decided to

rely on bulked seed samples (seed meal ground from ~5 g of seeds) in order to get a better average value without having to test a large number of samples.

The labour-intensive sample extraction methods, the cost of chemicals necessary for preparing samples for mass spectrometry (in particular the derivatisation agent and the internal standard) and limited availability of the instrument constrained the number of samples that could be analysed using the LCMS method using the internal standard. I therefore chose three mutant lines for detailed characterisation by measuring their ODAP concentrations in seedling shoots, seedling roots, leaves and roots, flowers, early and late pods and mature seeds. This range of tissues covered all developmental stages of the plant (seedling, juvenile, flowering, podding and seed maturity), in order to reveal any tissue-specific mutant phenotypes. I decided on using the mutant lines 4884-2 and 4946-7 because of their low ODAP concentrations measured by the spectrophotometric method (shown in Figure 44 in section 3.3.7). The mutant line 1874-11 showed the lowest seed ODAP concentration in the spectrophotometric experiment, but insufficient seeds of this line were available for it to be included in this experiment. I also included mutant line 1264-2, which showed intermediate ODAP-levels. For comparison, I included samples from all tissues of the varieties LSWT11 (the parent variety of the mutants) and Mahateora, the only low-ODAP variety to give consistent low-ODAP results in my experiments, as well as seed, seedling shoot and seedling root samples of pea (*Pisum sativum*) as a negative control. Three replicate extractions were made of each tissue. Samples were extracted, processed and measured according to the methods described in section 3.2.9.

Figure 50 shows the results of this experiment. The relative patterns of β-L-ODAP concentrations across the surveyed tissues were consistent among the five genotypes included in the experiment. The highest ODAP concentrations (2.8 % w/w of dry weight in LSWT11) were seen in the shoot tips of seedlings, followed by the root tips of seedlings. Lower levels of ODAP were observed in the leaves and flowers of the plant. ODAP concentrations in early pods of LSWT11 were notably higher than in late pods, consistent with earlier findings (Srivastava and Srivastava, 2006). This reduction in β-L-ODAP concentrations in developing pods may be caused by the β-L-ODAP being diluted by the large increase in dry mass during the seed filling stage. The ODAP concentration in LSWT11 seeds was 0.362 % w/w ± 0.007 % (standard error), a much higher value than measured using the spectrophotometric method, but in line with the LCMS measurements of single seed samples (which did not include the use of the SIL-internal standard). Only very small quantities of β-L-ODAP were found in the roots of the plant. The β-L-ODAP concentrations

in seedling shoots and seedling roots far exceed the β-L-ODAP concentrations in the seeds in all genotypes. This could suggest that β-L-ODAP is synthesised in both the shoot and root of grass pea seedlings at a high rate. However, these tissues have only a fraction of the dry mass of seeds. It is therefore possible that a large proportion of the β-L-ODAP observed in seedling tissues is relocated from the storage cotyledons during the germination process, despite β-L-ODAP in the seed being much lower as a percentage of dry weight than in the seedling shoot and root tips.
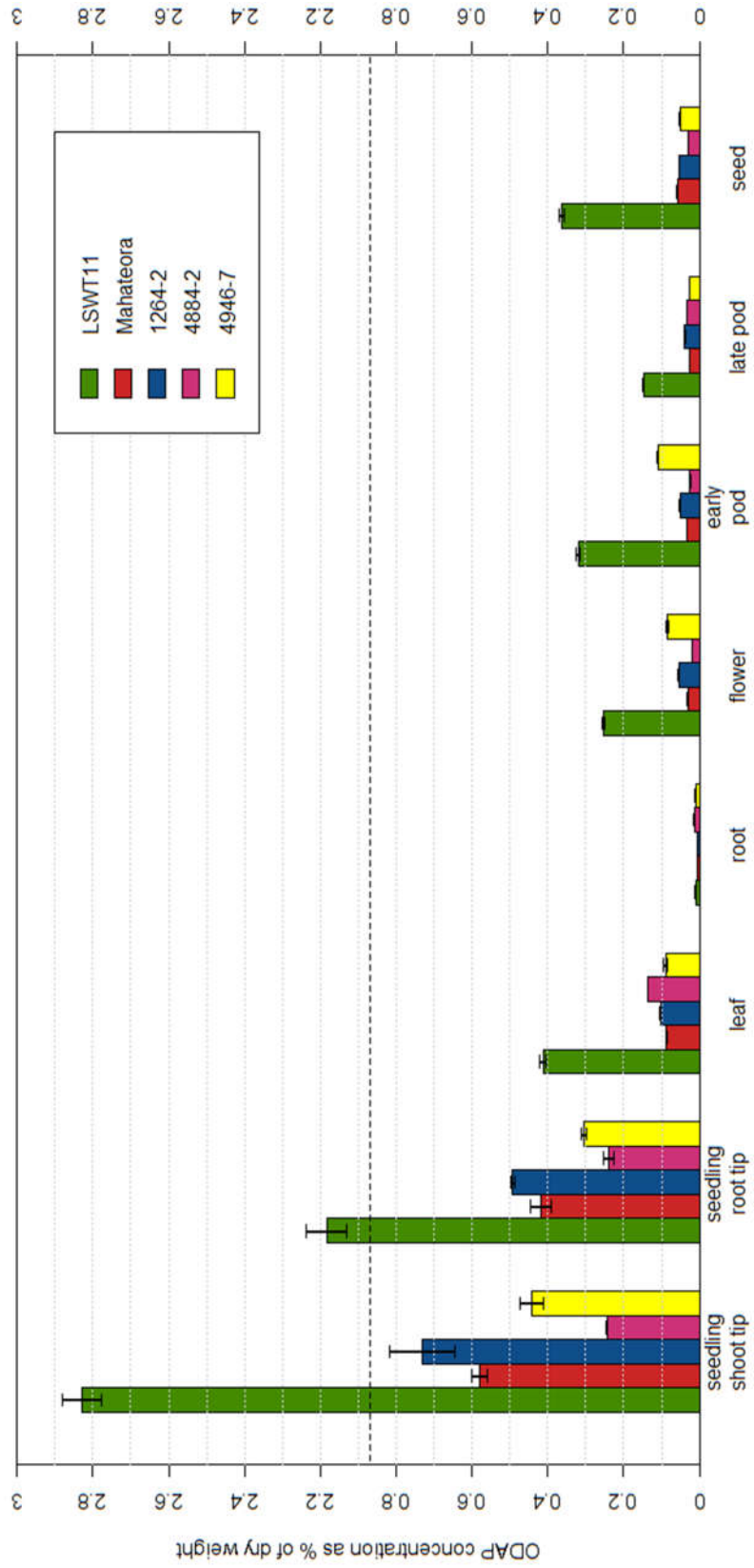
Figure 50. β-L-ODAP concentrations by tissue for five genotypes of grass pea, measured by LCMS using the SIL-internal standard. Error bars denote standard error of three replicate extractions.

The low-ODAP variety Mahateora and the three newly identified low-ODAP mutant lines each contained significantly lower amounts of β-L-ODAP in all tissues than LSWT11 (Student's t-test p<0.001), with the exception of roots, where the (extremely low) levels of β-L-ODAP were not significantly different between LSWT11 and the other genotypes. Pea (*Pisum sativum* cv. *Frisson*) seedling shoot tip, seedling root tip and seed meal samples, which were included in this experiment as a negative control did not contain any detectable amounts of β-L-ODAP. The mean seed β-L-ODAP concentration in Mahateora was 0.0581 % w/w ± 0.0007 % (standard error). The β-L-ODAP concentrations in all three surveyed mutant lines were significantly (p<0.01) lower than this. Lines 1264-2 and 4946-7 contained marginally less β-L-ODAP than Mahateora, with 0.0528 % w/w ± 0.0003 % and 0.0517 % w/w ± 0.0008 %, respectively. Seeds of the mutant line 4884-2, however, contained only 0.0285 % w/w ± 0.0006 % of β-L-ODAP, less than half the seed β-L-ODAP concentration of Mahateora. The mutant line 4884-2 also contained significantly lower amounts of β-L-ODAP in seedling shoots and seedling roots than Mahateora (p<0.01). These mutant lines may thus prove to be useful breeding material in developing grass pea varieties with further reduced β-L-ODAP content, especially if their low-ODAP phenotypes are due to mutations in different genes.

3.3.10 Accumulation of L-DAP in low-ODAP mutants

The spectrophotometric assay relies on the colour-forming reaction between L-DAP, β-mercaptoethanol and o-phthalaldehyde. In order to convert ODAP into DAP, the extracts were first hydrolysed. However, to exclude background noise caused by other compounds in the extract and the reading method, a second plate with non-hydrolysed samples was prepared and this reading was subtracted from the hydrolysed reading.

This conveniently served as a screening method for mutants accumulating L-DAP, which is believed to be the last intermediate in the synthesis of ODAP. In wild-type plants, L-DAP levels were below the sensitivity threshold of the spectrophotometric method. If any mutants were impaired in the last step of the synthesis, causing them to accumulate L-DAP, this should have produced high reading in the non-hydrolysed plate. However, no such samples were observed in the entire mutant screen. This points to the metabolic enzyme responsible for the last step of the synthesis of ODAP being redundant or, alternatively to a separate efficient breakdown mechanism for L-DAP that prevents it from accumulating in the young shoot.

A subset of samples used for the LCMS measurements to measure β-L-ODAP in the tissues of three mutant lines, LSWT11 and Mahateora was also qualitatively screened for the presence of O-acetyl-serine (OAS) and L-DAP. OAS was present in most tissues, but L-DAP was not detectable in any tissue of either the mutant lines or the varieties, despite the high sensitivity of the LCMS instrument.

3.3.11 Genetic analysis

In order to ascertain the number of separate mutated genes in a population of mutants with similar phenotypic alterations it was necessary to perform complementation tests to determine whether mutations were or were not allelic. This relied on the assumptions that the mutant phenotype (low/zero ODAP) was recessive and mutated function could be restored to its wild-type state by the presence of a single wild-type allele of each of the mutated genes. For this analysis, mutants were assumed to be homozygous, as they were selected from an M_2 population, and were crossed with each other. Where possible, each mutant in the set was crossed with each other mutant reciprocally. The F1 individuals resulting from these crosses were then screened for ODAP concentrations. Crosses in which the function was restored to the wild-type state could be expected to involve mutations in different genes, because the offspring carried one dominant functional allele of each mutated gene. Crosses in which the function was not restored could be expected to involve mutations in the same gene: in such cases the offspring carried two dysfunctional alleles, one from each of its parents. By performing this analysis on a set of mutants that were crossed in every combination, it was possible to infer the number of separate loci that were mutated in the set. If a sufficient number of mutants can be obtained from a mutant population, this can be used to derive the number of separate essential genes for a particular genetic function, such as the number of enzymatic steps in a metabolic pathway.

Two varieties of grass pea that are currently cultivated in India, Nirmal (Asthana, 1995) and Mahateora (Bhowmick, 2013), were included in the complementation analysis. Nirmal served as a high-β-L-ODAP parent line alongside the mutant parent variety LSWT11, while Mahateora was included to screen for complementation with the low-ODAP mutant lines. I decided not to include the low-ODAP variety Bio L212 (Ratan) in the complementation analysis, because the low-ODAP phenotype of Mahateora derives from Ratan. Hence, these two varieties are likely to be part of the same complementation group and only one needed to be included in the crossing. I decided to use Mahateora because its pink flower

phenotype allowed me to ascertain whether crosses with the blue-flowered LSWT11 plants and low-ODAP mutants had been successful.

I performed crosses between the low-toxin mutant lines 44-1, 51-9, 490-4, 992-8, 1172-1, 1264-2, 1390-4, 1548-3, 1874-11, 2003-7, 2158-10, 3720-9, 4884-2 and 4946-7. Because of the large number of lines in the crossing set, I aimed at performing each cross five times, including reciprocal crosses, with each line being the female parent in at least two of them. Due to poor performance of some of the plants, not all crosses could be completed. In total, I performed 671 individual crosses. Out of these, 250 crosses produced at least one seed, a take efficiency of 37.3 %. Crosses were performed in two batches.

To find out which mutants were allelic, I analysed the toxin concentrations of seeds resulting from these crosses and seeds from each parent line (mutant lines, LSWT11 and Nirmal) using the spectrophotometric assay. If a cross had produced more than one seed, I tested the ODAP concentrations in up to three seeds. All mutant lines and the variety Mahateora showed low concentrations of ODAP, while the parent variety LSWT11 and the variety Nirmal showed high toxin concentrations. After drilling into each seed to extract material for the assay, I allowed the seeds to germinate and planted them into individual pots. I harvested seedling shoots eight days after imbibition and again performed the spectrophotometric assay on these samples. The ODAP concentrations of the seeds resulting from the F1 crosses and the shoot tips of seedlings germinating from these seeds were strongly correlated, as shown in Figure 51.
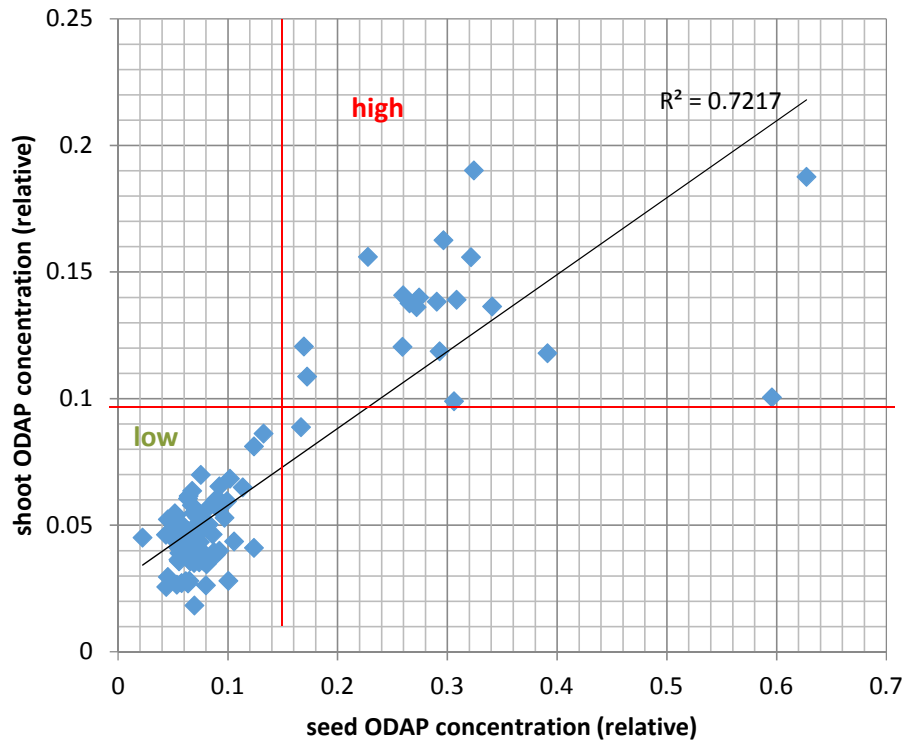
Figure 51. Correlation of ODAP concentrations in the seeds of high- and low-ODAP lines and the first batch of F1 hybrids with the ODAP concentrations in the shoot tips of seedlings which had germinated from these seeds. Red lines show the thresholds used to categorise individuals as high- or low-ODAP.

To identify events of gene complementation, I categorised the results of these measurements as either high- or low-ODAP, based on empirically assigned thresholds for the categorisation, as shown in Figure 51. The categorised crossing results are shown in Table 6.

Table 6. Categorised ODAP concentrations of parent lines and F1 hybrids resulting from the first batch of crosses, based on seed and seedling shoot results. In each case where more than one cross was successful and where more than one seed was produced by a cross, measurement resulted in the same categorisation. * indicates that two individual crosses of these parents were successful. Categorisations of parental lines are shown in italics

female parent	male parent	LSWT 11	44-1	51-9	490-2,3,4	992-8	1172-1	1548-2,3	1874-2,11	2003-6,7,10	Nirmal	Maha-teora
		high	*low*	*low*	*low*	*low*	*low*	*low*	*low*	*low*	*high*	*low*
LSWT11	*high*		high		high*	high*			high*		high	
44-1	*low*	low*										
51-9	*low*	low	low		low					low	low	
490-2,3,4	*low*	low								low		
992-8	*low*			low	low					low		low
1172-1	*low*											
1548-2,3	*low*		low			low*			low*	low	low	
1874-2,11	*low*	low		low	low	low				Low	low	high
2003-6,7,10	*low*	low			low*	low*					low	low
Nirmal	*high*	high										
Mahateora	*low*	low									low*	

Only a single cross between two low-ODAP lines (1874-11 x Mahateora), generated a high-ODAP F1 hybrid. All other crosses between low-ODAP parents resulted in low-ODAP F1 hybrids. All crosses between high-ODAP (LSWT11 or Nirmal) and low-ODAP parents (mutant lines or Mahateora) generated low-ODAP F1 hybrids if the female parent (pollen acceptor) was low-ODAP, but high-ODAP F1 hybrids if the male parent (pollen donor) was low-ODAP. These results were consistent between seeds and seedling shoot tips.

As shown in Table 6, reciprocal crosses between high- and low-ODAP parents generally did not result in F1 hybrids with the same ODAP concentrations. An example of this is shown in Figure 52. Both hybrids showed ODAP concentrations between their high-ODAP and low-ODAP parents, but the hybrid originating from the cross in which the maternal parent is low-ODAP contained much less ODAP than the hybrid of the reciprocal cross in which the maternal parent was high-ODAP.
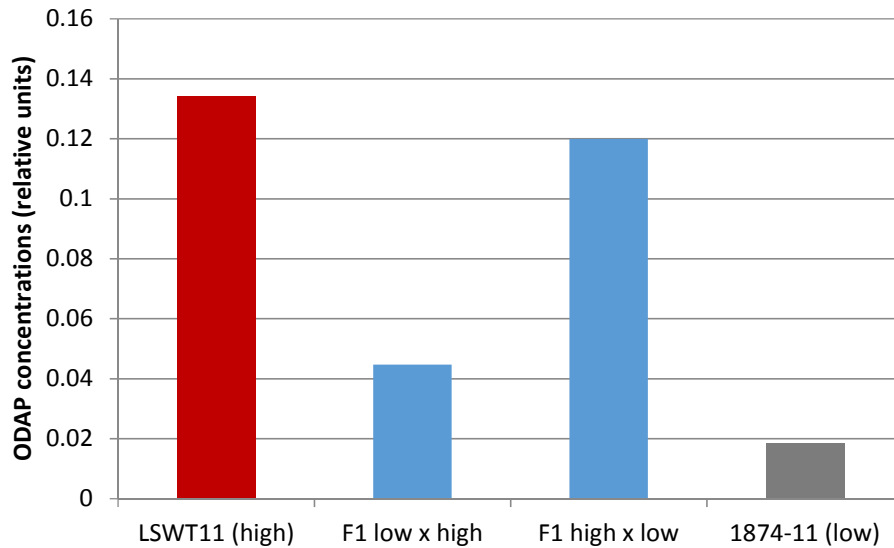
Figure 52. ODAP concentrations in seedling shoot tips for the reciprocal cross of LSWT11 with the low-ODAP mutant line 1874-11.

Plots showing the correlations between the ODAP concentrations in seedlings of F1 hybrids with the ODAP concentrations in their highest ODAP parent and in their maternal parents are presented in Figure 53. The correlation with the maternal parents was stronger (R^2 = 0.5862) than the correlation with the highest ODAP parent (R^2 = 0.1204). Correlations between F1 hybrids and the lower ODAP parent or the paternal parent gave R^2-values of 0.2354 and 0.0075, respectively (not shown in figure).
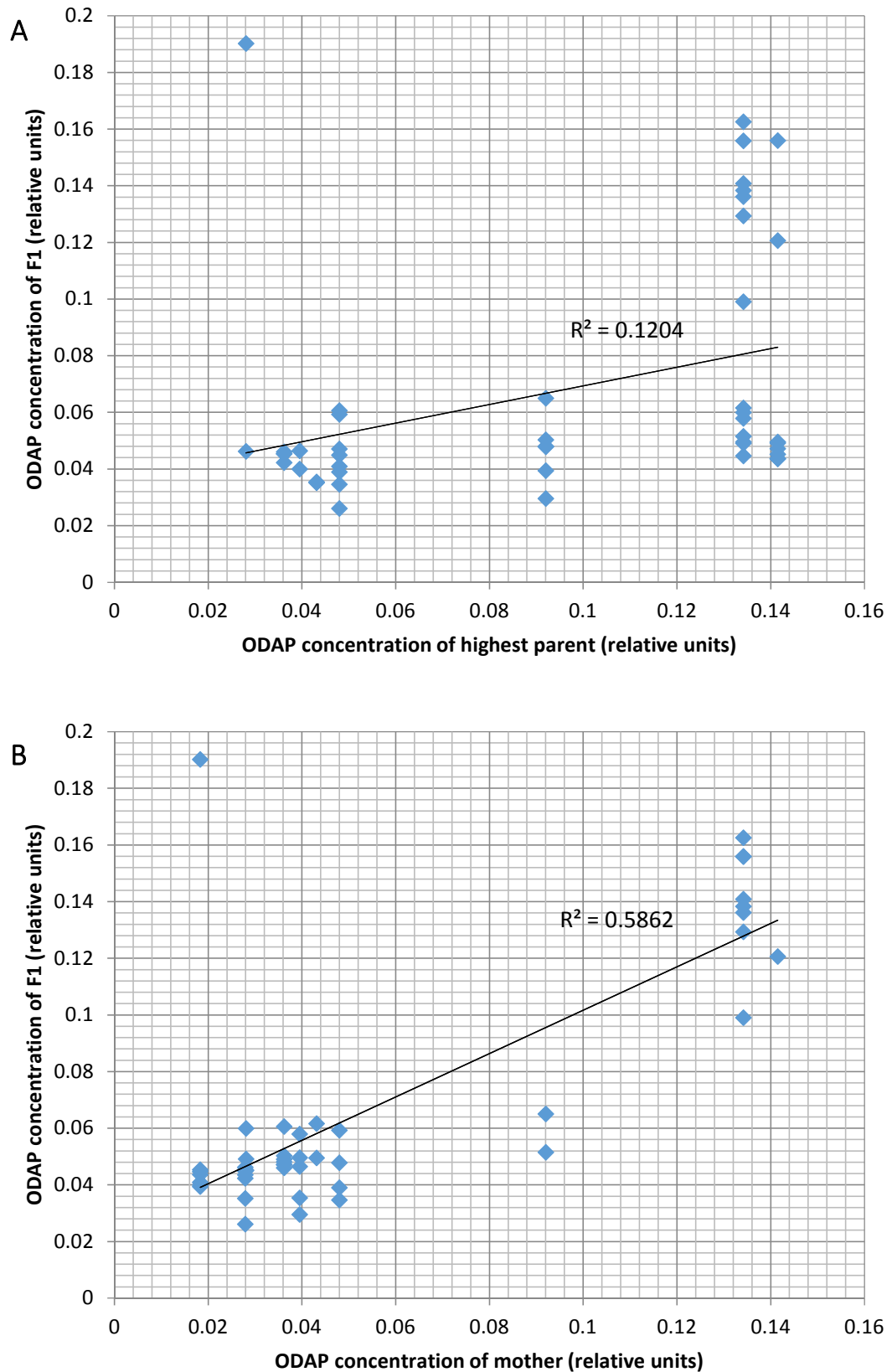
Figure 53. ODAP concentrations (relative units) in seeds resulting from F1 crosses between high- and low- ODAP plants plotted against the ODAP concentrations in seedling shoot tips of A) the line of their highest ODAP parent and B) the line of their mother plant. Linear trend lines are shown with their associated R² values.

If the high-ODAP phenotype were dominant, the ODAP concentration of the F1 hybrids would be determined by the parent with the higher ODAP concentration. However, ODAP concentrations of the seedling shoot tip of F1 hybrids were only poorly correlated with the higher ODAP parent. A better correlation existed with the ODAP concentrations of seedling shoot tips of the line used as the maternal parent.

There are several possible explanations for this apparent maternal effect. If a flower was not successfully pollinated during the crossing procedure and was not correctly emasculated, it may have self-fertilised. In this experiment, this would have resulted in offspring with ODAP concentrations similar to their maternal parents. This possibility could be excluded by monitoring other phenotypes in order to prove the crossing was successful. As none of these mutants differed from the parent variety in easily observable morphological phenotypes, this could not be done for the majority of crosses. However, the variety Mahateora produces pink flowers, a phenotype recessive to the blue flowers produced by LSWT11. Of the three crosses where Mahateora was the female parent and LSWT11 or a mutant line was the male parent, two produced plants with blue flowers, while one produced a plant with pink flowers. The pink-flowered plant was likely due to a self-fertilisation event and was excluded from the analysis, but the other two plants showed that low-toxin seeds were indeed produced from successful crosses between a low- and a high-toxin variety, but only if the low-toxin variety was the female parent.

The observed maternal effect in these two, and potentially other, crosses between high- and low-β-L-ODAP parents could be due to the deposition of ODAP into the developing seed by the maternal parent plant, i.e. a sporophytic depositional maternal effect. This β-L-ODAP could be relocalised into the developing shoot tip upon germination, leading to relatively high or low levels of β-L-ODAP in this tissue, depending on the mother plant. Similarly, if an intermediate of the biosynthesis downstream of the synthetic step that is disrupted by the mutation was deposited into the seed during its development, it could result in β-L-ODAP levels in the seed and seedling that were not representative of the plant itself, but of its maternal parent. While such an effect could result in a mutant phenotype in the embryo and the young seedling, it is unlikely to persist over the lifetime of the plant as any compound that is present in, but not produced by, the seedling would be diluted as the plant grows and potentially be degraded.

Alternatively, there may be a component of gametophytic maternal inheritance to the production of β-L-ODAP, either if an enzyme or regulatory factor involved in its biosynthesis is transferred cytoplasmically (if it is encoded by a chloroplastic gene), or if it is affected by genomic imprinting. This would be observed in the pattern of inheritance of high- and low-ODAP phenotypes, because the low-ODAP mutant phenotype would be passed down strictly in the maternal line and would persist throughout the lifetime of the plant. This would cause an inheritance pattern of the low-ODAP phenotype that looked as though all crosses had failed due to self-fertilisation.

There is disagreement in the literature about the inheritance pattern of β-L-ODAP content. While Nerkar (1972) described the inheritance of β-L-ODAP levels as Mendelian, despite observing continuous variation in toxin concentrations in the segregating generation, Quader et al. (1987) as well as Tiwari and Campbell (1996) described quantitative inheritance patterns when low- and high-ODAP varieties were crossed, with the F1 generation exhibiting intermediate levels and the F2 generation segregating continuously across the entire parental range. Comparing the results of crosses between four low toxin varieties, Tiwari and Campbell found the broad-sense heritability of ODAP concentrations in the seed to vary between 17 % and 92 %, indicating a complex pattern of control (Tiwari and Campbell, 1996). The maternal effect I observed was reminiscent of effects observed by Quader et al. (1987) and Tiwari and Campbell (1996) who speculated about the existence of cytoplasmic factors affecting ODAP content. However, no further evidence has been published to substantiate these hypotheses.

To test whether this maternal effect would persist as the F1-plants grew older, I repeated the analysis on top leaflets collected from the same plants four weeks later. The ODAP levels measured in these samples were lower than in seed and seedling shoot tip samples, while the background measurements (absorbance readings of non-hydrolysed samples) were much higher. This made categorisation of F1 hybrids as high- or low-ODAP difficult. For this reason, I introduced a 'medium-ODAP' category, which was not used to infer complementation. Categorisation of ODAP concentrations in leaf samples from batch 1 crosses are shown in Figure 54. The intermediate phenotypes seen in this experiment raise the possibility that ODAP concentrations in leaves follow a pattern of partial dominance, with heterozygotes displaying ODAP concentrations between the concentrations seen in homozygotic mutant and homozygotic wild type individuals.
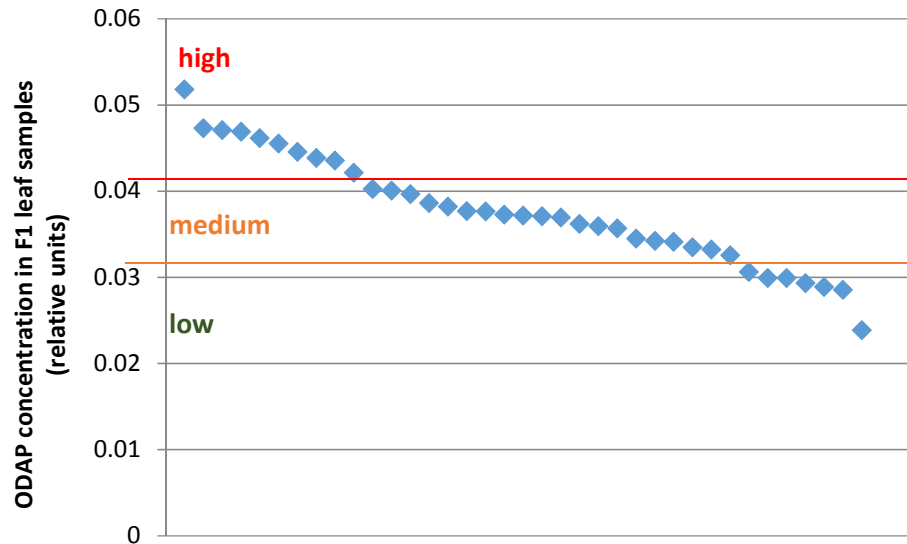
Figure 54. ODAP concentrations of leaf samples of five-week-old F1 plants from the first batch of crosses, ordered from high to low. Red and orange lines show the thresholds applied for categorisation into high-, medium- and low-ODAP individuals

As shown in Table 7, reciprocal crosses between high- and low-ODAP parents produced consistent results when leaf samples from 5-week-old plants were measured. All crosses between high- and low-ODAP parents resulted in F1 plants categorised as high- or medium-ODAP, regardless of whether the high-ODAP parent was maternal or paternal. This shows that the maternal effect observed in seedlings and seeds either does not affect leaf tissues or wears off as the plant grows. This would not be observed in plants that were the result of unwanted self-fertilisation. However, the possibility that those five-week-old 'F1' plants that still show the maternal phenotype in their leaves (e.g. the cross 992-8 x Mahateora in Table 7) have resulted from self-fertilisation cannot be excluded. This indicates that the maternal effect could be explained by sporophytic deposition or by gametophytic genomic imprinting, i.e. epigenetic marks that are lost as the plant ages. A gametophytic cytoplasmic effect (involving extra-chromosomally encoded genes) would not explain the wearing off of this effect.

Table 7. Categorised ODAP concentrations of parent lines and F1 hybrids resulting from the first batch of crosses, based on measurements of top leaflets from 5-week-old plants. Categorisations of parental lines shown in italics

Female parent	Male parent →	LSWT11	44-1	51-9	490-2,3,4	992-8	1172-1	1548-2,3	1874-2,11	2003-6,7,10	Nirmal	Maha-teora
		high	*low*	*low*	*med*	*low*	*low*	*low*	*low*	*low*	*high*	*low*
LSWT11	*high*	—	high		high	high			high		high	
44-1	*low*	high	—									
51-9	*low*	high	med	—	med					high	med	
490-2,3,4	*med*	med			—					med		
992-8	*low*			med	low	—				low		low
1172-1	*low*						—					
1548-2,3	*low*		med			med		—		med	med	low
1874-2,11	*low*	high		med	med	high			—	med	med	med
2003-6,7,10	*low*	high			low	med				—	med	low
Nirmal	*high*	high	high								—	
Maha-teora	*low*	med									low	—

However, the very low overall level of ODAP in these leaf samples and the variability of the background reduced the number of data points that could be used to infer complementation groups. I therefore analysed additional crosses. Because of the high degree of correlation between seed and seedling shoot ODAP concentrations, I skipped the collection of seed material by drilling each seed and only harvested seedling shoot samples for analysis. As with the first batch of crosses, the ODAP concentration of F1 seedlings resulting from crosses between high- and low-ODAP parents corresponded to the ODAP concentrations in their maternal parents (not shown). Hence, these data were not used to infer complementation. Leaf samples were collected from eight-week-old plants. To gather additional data, both the youngest pair of leaflets and the youngest fully expanded leaf were collected and assayed separately. Individuals were categorised as high-, medium- or low-ODAP based on both measurements as shown in Figure 55. Individuals for which the two measurements were in disagreement were not included in the inference of complementation groups. When processing these samples, I observed much higher readings from non-hydrolysed samples than previously seen. Subtracting these measurements from the readings of hydrolysed samples led to negative values in some cases and potentially contributed to the issues of disagreement between youngest leaflet

and expanded leaflet samples. This shows the limitations of the spectrophotometric assay when measuring ODAP in adult green tissues.
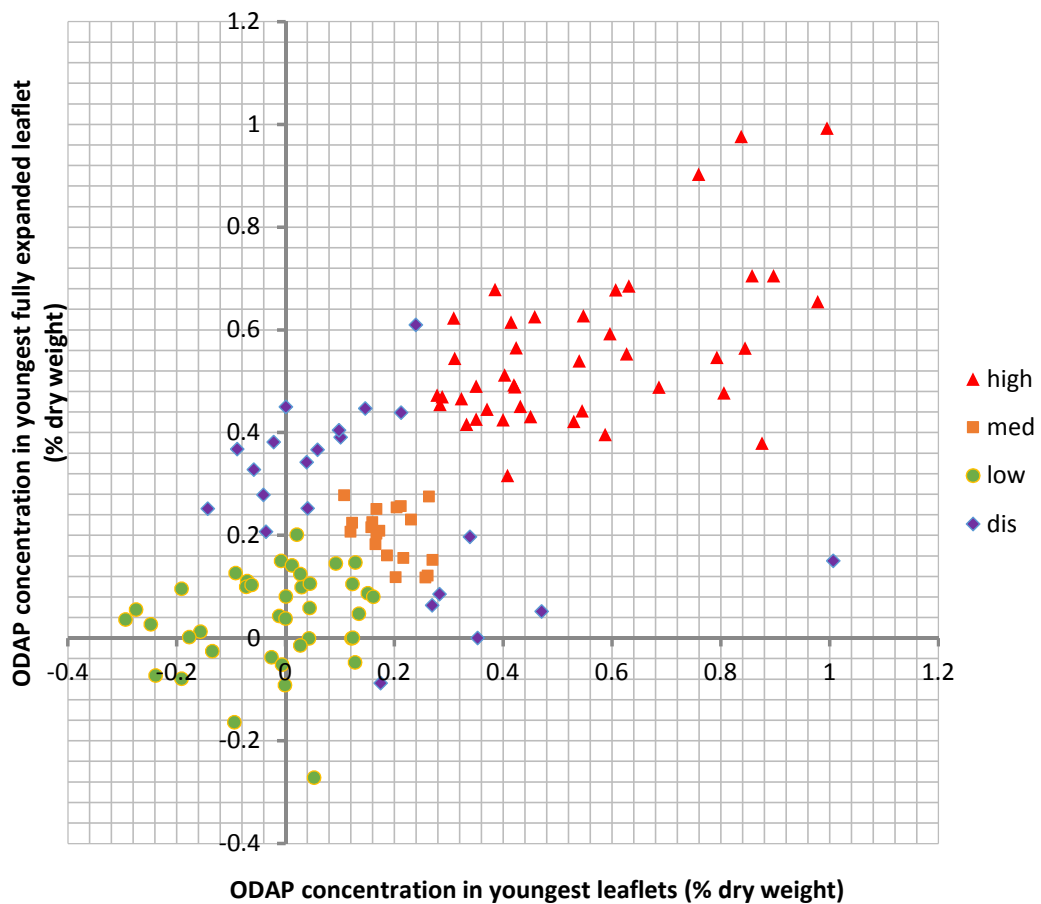
Figure 55. ODAP concentrations of leaf samples of eight-week-old F1 plants of the second batch of crosses. ODAP concentrations of the youngest leaflets are plotted against concentrations of fully expanded leaflets of the same plants. Symbols in red, orange and green show F1 plants categorised as high-, medium- and low-ODAP, respectively. Individuals for which the two samples disagreed are shown in purple. Negative values result from high background readings.

Despite these caveats, several crosses between low-ODAP mutant lines resulted in high-ODAP F1 plants, indicating complementation. I proceeded to infer complementation groups based on high- and low-ODAP hybrids, assuming that crosses of low-ODAP mutant in the same complementation group would result in a low-ODAP hybrid, while crosses between mutants in different complementation groups would result in high-ODAP hybrids. The combined results of leaf measurements of both batches of crosses and the hypothesised complementation groups are shown in Table 8. Two complementation groups emerged from these data. Group 1 includes the mutant lines 44-1, 992-8, 1264-2, 1548-3, 1390-4, 2003-7 and 2158-10 as well as the low-ODAP variety Mahateora. Group 2 includes the mutant lines 490-2, 1874-11, 3720-9 and 4946-7. Some crosses resulted in conflicting results. For example, the cross 2003-7 x 51-9 produced a low-ODAP hybrid, while the

reciprocal cross (51-9 x 2003-7) resulted in a high-ODAP hybrid. Low ODAP readings in which the phenotype of the supposed hybrid matches the mother plant may be the result of failed crosses (i.e. self-fertilisation events). This is likely the case in the Mahateora x Nirmal cross that produced a low-ODAP plant. For this reason, crosses between low-ODAP parents that apparently resulted in low-ODAP F1 offspring could not be counted as conclusive evidence that both parents belong to the same complementation group. Hence, the complementation hypothesis presented here is preliminary and will be confirmed or revised using ODAP concentration data from seeds produced by the F1 plants (ongoing). Three mutant lines, 51-9, 1172-1 and 4884-2 could not be assigned to either group and may be members of additional complementation groups. This places two of the mutant lines which were characterised by LCMS (1264-2 and 4946-7), as described in the previous section, into separate complementation groups, while line 4884-2 cannot be assigned to either group based on the available evidence.

Table 8. Categorised ODAP concentrations of leaves of F1 hybrids produced by both batches of crosses, organised by hypothesised complementation groups.

	varieties		group 1								group 2				no group		
	LSWT11	Nirmal	44-1	992-8	1264-2	1390-4	1548-3	2003-7	2158-10	Maha-teora	4946-7	490-2	3720-9	1874-11	1172-1	51-9	4884-2
varieties LSWT11		high	high	high	high		high	high	high	med		med		high	high	high	high
Nirmal	high		high		high		med	med	high	low		dis	dis	med	high	high	high
group 1 44-1	high	high		dis	dis		med	low	low	dis	low	low	high	high			
992-8	high		dis			low	low	low					low			med	med
1264-2	high	high	dis			low	med			low		high			high		
1390-4				low	low					med	med	high	dis	low			
1548-3	high	med	med	low	med			med			high	dis		low		low	low
2003-7	high	med	low	low			med				med	dis	low			low	
2158-10	high	high	low								high		low		dis	low	
Maha-teora	med	low	dis		low	med						low	low	low			high
group 2 4946-7			low			med	high	med	high					med	high		high
490-2	med	dis	low		high	high	dis	dis		low			dis	med		med	med
3720-9		dis	high	low		dis		low	low	low		dis					med
1874-11	high	med	high			low	low			low	med	med				med	med
no group 1172-1	high	high			high				dis		high						
51-9	high	high		med			low	low	low			med		med			med
4884-2	high	high		med			low			high	high	med	med	med		med	

The observation of gene complementation in these crosses showed that not all of the observed mutations are in the same genes. The complementation hypothesis compiled from these data implied that at least two mutated genes are represented among the 15 low-ODAP lines included in the crossing. The three ungrouped mutant lines may be placed in either group with more evidence or they may form one or more additional groups. These mutants may thus provide a useful tool for understanding the genetics of β-L-ODAP production in grass pea. However, many crosses had to be excluded from the complementation analysis, because the resulting individuals could not be categorised as high- or low-ODAP. To gain more confidence in these complementation groups, the analysis should be repeated using seed material produced by the F1 plants, which I was not able to complete due to time constraints. These seeds themselves are F2 individuals, i.e. the selfed offspring of the F1 hybrids. The ODAP concentration in the seeds can be expected to be subject to the maternal effect observed previously. As ODAP measurements in seed samples did not suffer from the high background readings observed in adult green tissues, these results should allow a better categorisation of samples into high- and low-ODAP, enabling complementation groups to be inferred with more confidence.

Importantly, the complementation observed in the crossing experiments with Mahateora indicate that some of the mutants are likely to contain mutations in different genes involved in the synthesis of β-L-ODAP or its regulation than the genes that confer the low-ODAP phenotype in this variety, which is related to several other low-ODAP grass pea varieties that are currently being cultivated. Some of the mutants described here thus appear to contain mutant alleles that could be introgressed into existing low-ODAP varieties to further reduce their ODAP contents.

3.3.12 Other mutant phenotypes

While the content of ODAP in the young shoots was the only phenotype for which the entire mutant population was screened methodically, a number of other mutations also became apparent in the population. Among these were chlorophyll mutations such as *albino*, *chlorina* and *xantha* mutations that became visible in the young seedlings. While the numbers of these can give some indication of the mutation density in the population, there was little value to these mutants otherwise, as none of them survived through to seed set.

While the collected shoot samples were screened for their ODAP content, the plants were left in the trays for more than six weeks. This was enough time for many of them to start

producing flowers. Several families included plants with changed flower colours. Besides the wild-type purplish blue flowers, with pink spots at the centre of the standard petal, the following mutant phenotypes were observed: white flowers with blue centres, entirely pink or peach-coloured flowers and entirely white flowers. These mutant phenotypes are shown in Figure 56.

Figure 56. Types of flower colour mutants observed in the mutant population, showing back and front views. A,B) wild type; C,D) partial blue; E,F) peach; G,H) white

White flowers with pink centres, which are seen in some Indian landraces were not observed, neither were entirely blue flowers without the pink centre or back. This implied that the synthesis of the flower pigments (which are anthocyanins) proceeds from the pink to the blue compound, as no mutation led to the loss of the pink, but not also of the blue pigment. The observed mutant flower colours may be interesting to breeders of ornamentals as well as food crops, as flower colour is an easily observable phenotype that can be used to differentiate a released variety from other varieties and landraces. Since these mutant phenotypes are recessive to the wild-type blue-flowered phenotype, changes in flower colour could also serve as a warning that a population of plants has been contaminated by outcrossing.

3.4 Summary

By adapting an existing spectrophotometric method for measuring ODAP content in grass pea tissues, I developed a high-throughput assay enabling a single researcher to screen over a thousand samples in one day. Using this assay I screened 3060 M2 families of a EMS-mutagenised grass pea population, a total of 36696 plants for mutants with reduced ODAP content. A total of 41 M2 families were selected in three passes of the screening method. The selected mutants were confirmed through re-testing in the following generation. Several of these mutants contained substantially lower amounts of ODAP than their parent variety and would be classed as 'low-toxin varieties' according to the widely used threshold of 0.1 % w/w of β-L-ODAP in dry seeds, but no mutant with zero ODAP content was identified. To confirm these results using an independent method, I synthesised β-L-ODAP labelled with carbon-13 atoms to serve as an internal standard for LCMS. Using an LCMS method developed for this purpose I further characterised three mutant lines, designated 1264-2, 4884-2 and 4946-7 by measuring β-L-ODAP concentrations in seedling shoot and root tips, leaves, roots, flowers, early and late pods and mature seeds. Tissues of the parent variety of the mutant population, LSWT11, and the low-ODAP variety Mahateora were included for comparison. Comparing samples grown in the same environment and measured using LCMS, all three mutant lines showed strongly reduced β-L-ODAP concentrations in all tissues except roots compared to LSWT11, and lower seed β-L-ODAP concentrations than Mahateora. The mutant line 4884-2 showed a seed β-L-ODAP concentration less than half that of Mahateora. Fourteen low-ODAP mutant lines selected from the screen and Mahateora were included in a complementation screen. This analysis was complicated by an apparent maternal effect on ODAP concentrations in seed and seedling tissues and high background readings when testing older tissues. However, the data showed complementation between some of the low-ODAP mutants, indicating at least two separate allelic groups. The newly identified mutants will form very valuable breeding material for the development of future low-ODAP varieties of grass pea.

Chapter 4 – Identification of candidate genes encoding metabolic enzymes in the β-L-ODAP biosynthetic pathway

4.1 Introduction

4.1.1 A reverse genetics approach to identify target genes for the development of zero-toxin grass pea genotypes

The low-toxin mutants that I was able to identify, described in the previous chapter, shared two important characteristics with the low-toxin varieties of grass pea that had previously been released: they produced low, but measurable non-zero levels of β-L-ODAP. Reports exist of previously described low-ODAP varieties with markedly increasing toxin contents under stressful conditions (Fikre et al., 2011; Fikre et al., 2006; Girma and Korbu, 2012). My objective was to develop zero-toxin lines, because unless toxin content can be shown to be reliably below an as yet-to-be-established safe threshold consistently, it will be difficult to establish that any new mutant plants are safer than existing varieties. Moreover, the intermediate levels of toxin resulting from high-low crosses suggest a strong maternal influence on ODAP levels. This means that a better understanding of the genetic basis of the toxin production, in particular of the enzymes involved in the synthesis, was necessary. Identifying genes encoding enzymes responsible for toxin synthesis, would open the way to developing grass pea plants unable to produce β-L-ODAP through targeted screening for mutations in these genes using TILLING (Henikoff et al., 2004) or disruption of those genes by means of genome editing (Belhaj et al., 2013).

One way to identify such genes is to use a forward genetics mapping approach, including crossing high- and low-ODAP lines that differ in a number of genetic markers and measuring the toxin levels in the F2 offspring of these crosses to map genes involved in toxin production. However, this would require a genetic map for grass pea with sufficient markers to map genes with a high degree of accuracy, which does not currently exist (Skiba et al., 2007). While a large set of microsatellite loci has recently been described in grass pea (Yang et al., 2014), only a small minority of these microsatellites have been characterised as

polymorphic markers or linked to a genetic map of *Lathyrus sativus* chromosomes (Lioi and Galasso, 2013; Sun et al., 2012). In addition, these experiments would be complicated by the long generation time of grass pea, the difficulty of reliably crossing its self-pollinating flowers and the non-trivial problems with scoring ODAP-production phenotypes. While such an approach might identify important quantitative trait loci (QTL), it would be unclear, until the very end, whether such QTL represent regulatory or metabolic genes and whether they could be useful targets for developing entirely toxin-free plants. While a forward genetics approach to mapping and identifying biosynthetic genes is feasible, it would require a considerable amount of time and resources.

A reverse genetics approach might be considered more promising. By gathering a genome-wide dataset of expressed genes, it should be possible to identify candidates for the genes encoding enzymes of the pathway given the existing knowledge of the β-L-ODAP biosynthetic pathway. A limited number of candidate genes could be followed up experimentally to confirm whether any encode enzymes that catalyse the relevant reactions. These genes would provide very promising targets for reverse genetics approaches to develop reduced-toxin or toxin-free grass peas.

No suitable genome or transcriptome dataset existed for grass pea that could be used to identify candidate genes. The genus *Lathyrus* contains species with a wide variety in genome sizes from 3.35 Gbp up to 16.6 Gbp (Narayan, 1998). The genome size of *Lathyrus sativus* appears to be subject to intraspecific variation (Ghasem et al., 2011; Nandini et al., 1997), leading to the genome size of various *Lathyrus sativus* accessions being estimated as low as 6.8 Gbp and as high as 10.8 Gbp (Nandini et al., 1997). Despite the rapid advances in genome sequencing technology in recent years, sequencing and assembling a genome of this size is a significant undertaking.

The grass pea genome contains a large amount of repetitive sequences and many potentially duplicated gene regions (Narayan, 1982), making the assembly of reads into large contiguous sequences likely to be difficult. Hence, producing an accurate and reliable reference genome for grass pea from which candidate genes could be extracted is likely to be time consuming and very expensive. No such genome sequencing project has been started to date. The African Orphan Crops Consortium which, in 2014, set out to sequence the full genomes of 101 plant species (Fox, 2013), decided to forego grass pea, despite including species of far lower economic importance, such as baobab, species with much

larger genomes such as fava bean, and species that can hardly be classified as African orphan crops, such as onion.

Rather than sequencing the entire genome of grass pea, a more affordable, and for the purpose of finding candidate genes perhaps more promising approach, involves sequencing the transcriptome of grass pea followed by de-novo assembly without a reference genome (Grabherr et al., 2011). This approach reduces the overall amount of sequencing that is necessary, as only the expressed genes are sequenced, while the much larger non-coding regions of the genome are ignored. This also reduces the difficulty in assembling the reads computationally, as the coding sequences are likely to be less repetitive than non-coding sequences. Sequencing the transcriptome nevertheless raises some complications, as it captures only genes that are currently expressed in the tissues from which RNA is extracted. Genes that are only expressed at very low levels, for example transcription factors, and secondly genes that are only highly expressed in specific tissues or developmental stages may be missed or mis-assembled due to insufficient read depth. In trying to find candidates for the metabolic enzymes involved in the synthesis of β-L-ODAP, the low expression might not be a problem in this case, as the enzymes necessary for the synthesis of such a common metabolite (>2 % of dry weight in some tissues) are likely to be expressed at significant levels. Even if there are copies of these genes with low expression that play a minor role in the synthesis, it should be possible to identify these by homology searches, if the major genes can be found first. The second issue of missing genes because they are primarily expressed in other tissues might pose more of problem, because the exact sites of β-L-ODAP synthesis are unknown. While the toxin accumulates in the young shoot tips, young leaves, immature pods and seeds as shown in the previous chapter (see section 3.3.9), it is unclear whether toxin synthesis is restricted to these tissues or whether some of the toxin is transported to these tissues from elsewhere. This means that RNA from several tissues, representing different parts of the plant and different developmental stages needs to be sequenced in order to ensure capture of the relevant genes and to be able to compare their expression levels between tissues.

Grass pea was one of the species included in the 1000 Plant Genomes Project and an RNA sample was sequenced (Matasci et al., 2014) (despite its title, this project involved the sequencing of transcriptomes, not whole genomes), but the tissue that was used in this experiment (mature roots) does not contain significant amounts of ODAP and is therefore unlikely to be a major site of ODAP production. This means that genes encoding enzymes that are responsible for the synthesis of β-L-ODAP in high-toxin tissues (the germinating

shoot tip, young leaves and early pods) might not be expressed in the roots and could be missed as candidates.

In 2015, another transcriptome dataset from grass pea was published (Chapman, 2015), however, this dataset also covered only one tissue (leaves) and is based on a small number of reads (15.8 million). In order to identify candidate genes encoding enzymes involved in β-L-ODAP synthesis, a more extensive transcriptome, covering high- and low-toxin tissues was needed.

4.1.2 The putative pathway for β-L-ODAP synthesis suggests candidate gene families for some of the enzymes involved

The biosynthetic pathway leading to the production of β-L-ODAP is still poorly understood. The currently proposed biosynthetic pathway is shown in Figure 57. By feeding grass pea seedlings with radiolabelled β-(isoxazolin-5-on-2-yl)-alanine (BIA), Kuo et al. showed that BIA is an intermediate in the synthesis of β-L-ODAP (Kuo and Lambein, 1991; 1994; 1998). BIA itself is produced by many legume species, including pea, lentil and chickpea (Kuo et al., 1998). Ikegami et al. were able to show that BIA is converted to L-DAP, the immediate precursor of β-L-ODAP (Malathi et al., 1970) in vitro by an enzyme extracted from grass pea (Ikegami et al., 1999), but the reaction mechanism and the exact intermediate steps remained unclear.

In a separate reaction, oxalyl-CoA is formed by the ligation of oxalic acid and coenzyme A (Malathi et al., 1970). The enzyme catalysing this reaction may be a homolog of the recently identified oxalyl-CoA synthases from *Arabidopsis thaliana* (Foster et al., 2012) and *Medicago truncatula* (Foster et al., 2016). Oxalyl-CoA then acts as the acyl-donor for the acyltransfer reaction forming β-L-ODAP. Oxalyl-CoA is part of a pathway for the breakdown of oxalic acid in plants. Oxalic acid is a common metabolite in plants, and is involved in several catabolic pathways (Yu et al., 2010), but also serves a variety of functions itself, including metal detoxification and deterrence against insect feeding (Franceschi and Nakata, 2005). However, oxalic acid is also produced by some necrotrophic pathogens, in particular *Sclerotinia sclerotiorum* (Williams et al., 2011) as a means to suppressing host defences and inducing cell death (Kim et al., 2008).
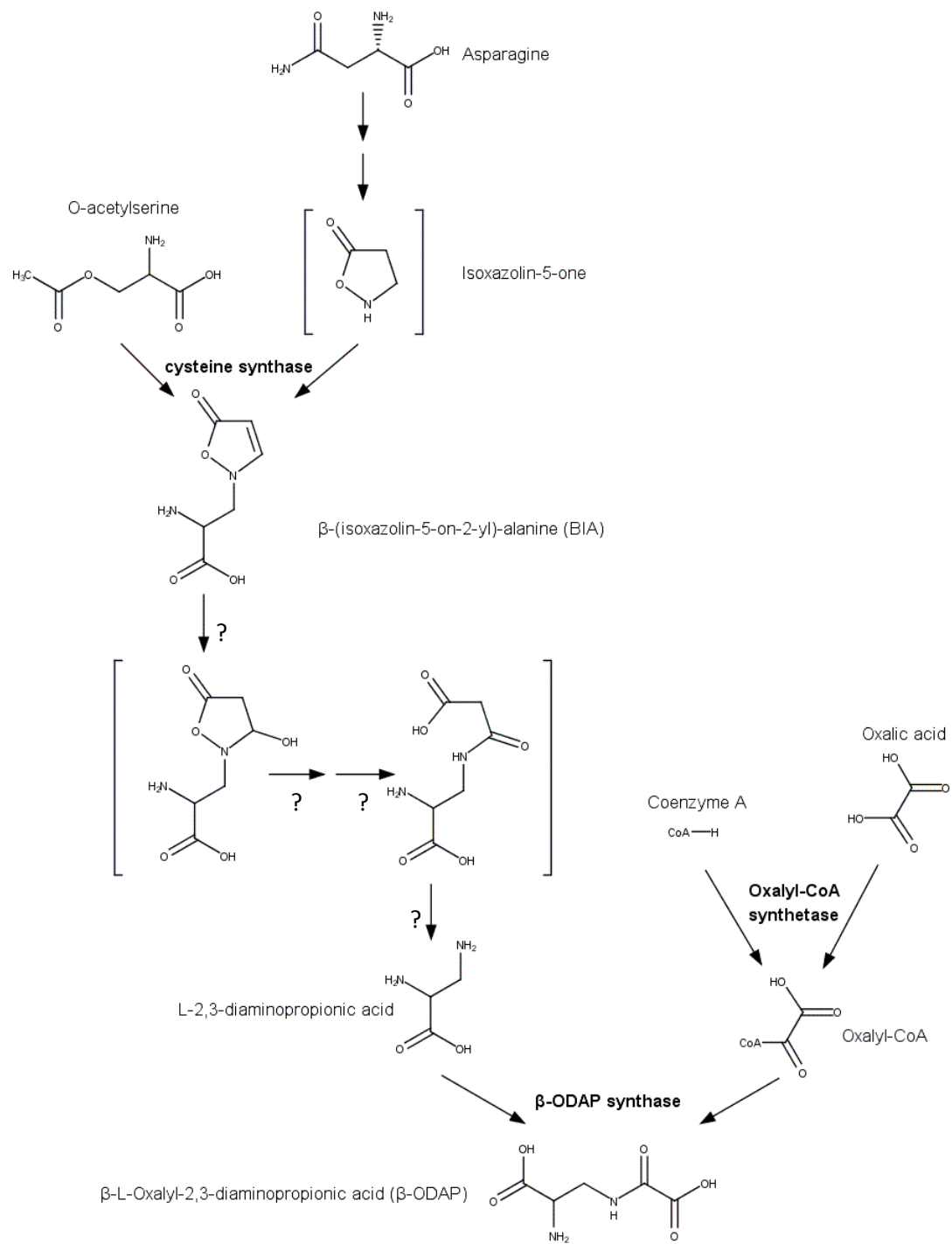
Figure 57. Proposed pathway for β-L-ODAP synthesis. Redrawn based on publications by Yan et al., Malathi et al., Kuo, Ikegami and Lambein (Malathi et al., 1970; Ikegami et al., 1991; Ikegami et al., 1999; Kuo and Lambein, 1991; Kuo et al., 1994; Kuo et al., 1998; Yan et al., 2006). Hypothetical intermediates are shown in square brackets. None of the proposed enzymes (shown in bold) have been identified to date in grass pea. (Note: this figure is identical to Figure 4 in Chapter 1 and has been reproduced here for ease of reference.)

In order to prevent oxalic acid (and calcium oxalate crystals) from accumulating to toxic levels, many plants have pathways for breaking down oxalic acid (Lane et al., 1993; Foster et al., 2012). One of these pathways converts oxalate to oxalyl-CoA, which is then decarboxylated to form formyl-CoA. Enzymes catalysing this reaction have been partially purified from grass pea (Malathi et al., 1970) as well as several other legume species (Adsule and Barat, 1977), but no corresponding gene has yet been identified in these species. However, an enzyme catalysing the formation of oxalyl-CoA was recently identified in *Arabidopsis thaliana* (Foster et al., 2012) and named *ACYL ACTIVATING ENZYME3 (AtAAE3; At3g48990).* Homologs of this enzyme that are able to catalyse the same reaction have since been identified in the budding yeast *Saccharomyces cerevisiae* (Foster and Nakata, 2014) and in the model legume *Medicago truncatula* (Foster et al., 2016). Green fluorescent protein (GFP) tagging has revealed that this enzyme is most likely localised in the cytosol (Foster et al., 2012; 2016). The *Medicago truncatula* enzyme MtAAE3 consists of 515 amino acids and is encoded by the open reading frame (ORF) of the gene Medtr3g035130. The sequence of both the *A. thaliana* and the *M. truncatula* enzymes could be used to identify oxalyl-CoA synthetase candidate genes from the grass pea transcriptome through sequence homology.

The final reaction, forming β-L-ODAP from L-DAP and oxalyl-CoA, has been suggested to involve an acyltransfer reaction. Two plant enzyme families have been described which catalyse acyltransfer reactions transferring groups other than aminoacyl-groups: the Serine Carboxypeptidase-Like (SCPL) acyltransferases and the BAHD acyltransferases (Bontpart et al., 2015), a family named after its first four enzymes to be identified: benzylalcohol O-acetyltransferase (BEAT), anthocyanin O-hydroxycinnamoyltransferase (AHCT), anthranilate N-hydroxycinnamoyl/benzoyltransferase (HCBT), and deacetylvindoline 4-O-acetyltransferase (DAT).

SCPL acyltransferases belong to a large family of enzymes which catalyse acyltransfer reactions using donor 1-O-glucosides (Milkowski and Strack, 2004). It is unlikely that the ODAP-forming reaction is catalysed by an enzyme of this family, as it would not use oxalyl-CoA, but would instead require Oxalyl-1-O-glucoside, contrary to the existing experimental evidence (Malathi et al., 1970). In addition, all SCPL-acyltransferases that have been described to date catalyse O-acylations, rather than N-acylations (Bontpart et al., 2015). BAHD acyltransferases, on the other hand, use CoA-derived donor molecules and are known to catalyse both O-acylations and N-acylations. This highly versatile superfamily of enzymes facilitates a wide range of metabolic steps in plants involving the transfer of acyl

groups, such as the acylation of anthocyanins (Unno et al., 2007), the production of green-leaf volatiles for herbivory defence (D'Auria et al., 2007a), the production of floral scent compounds (Dudareva et al., 1998), the esterification of cell wall saccharides (Bartley et al., 2013) and many other functions. Overall, more than 60 BAHD-acyltransferases have been assigned biochemical functions to date (Tuominen et al., 2011).

To identify candidate genes for the BAHD-acyltransferases catalysing the synthesis of β-L-ODAP in grass pea, I posited four criteria that these genes are likely to fulfil:

(i)   The gene either does not have a close homologue in related species that do not produce ODAP, or differs in the predicted amino acid sequence of the metabolically active HXXXD-domain
(ii)   The gene is part of a cluster of closely related genes in grass pea
(iii)   The gene has homology to known enzymes that catalyse N-acylation reactions
(iv)   The gene's pattern of expression across tissues and developmental stages of the grass pea plant is correlated with the pattern of toxin distribution

The BAHD-acyltransferase (BAHD-AT) family of genes is well known for its biochemical versatility: individual enzymes may accept several substrates, catalysing different reactions, and small changes in amino acid sequence may or may not result in major changes in substrate specificity and biochemical function (D'Auria, 2006; Ma et al., 2005). This makes it difficult to draw conclusions regarding the likely functions of a BAHD-AT gene based on its sequence alone. The highly conserved HXXXD-domain, which determines the shape of the enzyme's binding pocket, is particularly sensitive to changes in sequence. Mutations of the conserved histidine and aspartic acid residues generally result in non-functional enzymes, though some known BAHD-ATs do not share the glycine that typically follows this sequence (D'Auria, 2006). Mutations among the three variable amino acids in the domain can easily alter substrate specificity and this may be a mechanism by which BAHD-AT family members with novel functions evolve, making enzymes with altered active sites the most promising candidates for identifying new functions.

4.1.3   Transient expression allows assays of gene function *in planta*

To establish whether a candidate gene fulfils a certain biological role, one of the most informative and straightforward methods is to express it in another organism and test whether the function of interest is gained by the host (Kapila et al., 1997). In plant systems the expression of foreign genes can be achieved through stable transformation or through transient expression (Marillonnet et al., 2005). Stable transformation requires the germline of an organism to be transformed with the foreign DNA, allowing it to be passed on to the

offspring. This is particularly important when studying the effect of genes at different stages of plant development or when the expression of the transgene in future generations of the plant is desired, such as in the development of a transgenic crop. Functional studies using transient expression, on the other hand, deliver results much more quickly, as the gene of interest can be delivered into the cells of an already established plant, allowing its effects to be measured a few days later. This also allows genes to be studied that might have detrimental effects if they were expressed at a high level at earlier stages of development.

For testing the ODAP synthase candidates, transient expression was chosen, because it allowed results to be gained rapidly. For transient expression experiments, I decided to use *Nicotiana benthamiana* as a host plant, because it grows quickly, is easily infiltrated and can deliver very high levels of transient expression of foreign genes (Sainsbury and Lomonossoff, 2008). Crucially, *N. benthamiana* does not produce β-L-ODAP naturally, allowing me to test for this function in a β-L-ODAP-free background. The pEAQ-HT vector system developed by the lab of George Lomonossoff is specifically adapted to produce very high levels of expression of transgenes in *Nicotiana benthamiana* (Sainsbury et al., 2009; Peyret and Lomonossoff, 2013) and has been used in many studies using transient expression assays (Kanagarajan et al., 2012; Duvenage et al., 2013; Tan et al., 2014). It can be used in conjunction with the Gateway cloning system (Invitrogen) to rapidly produce highly efficient expression vectors containing any ORF of interest.

As described in the previous chapter (see section 3.3.9), LCMS provided a highly sensitive method of measuring the presence of β-L-ODAP in plant tissues. This has allowed me to measure the production of small amounts of β-L-ODAP in *N. benthamiana* tissues expressing candidate genes from grass pea to identify ODAP-forming enzymes.

4.2 Materials and methods

4.2.1 Extraction of RNA from grass pea tissues

Samples were collected from seven tissues of grass pea (accession LSWT11) plants. These tissues were the same as those used for mass spectrometric measurements of β-L-ODAP, described in Chapter 3 (see section 3.2.9), with the exception of mature seeds, which were not included in the transcriptome. The tissues were collected and ground in the same way as described in section 3.2.9. Briefly, tissue samples were collected from the shoot and root tips of seedlings 8 days after imbibition, from whole leaves and whole roots of 5-week-old plants and flowers, and from early and late pods of 2-month-old plants. The samples were flash-frozen in liquid nitrogen and ground using a mortar and pestle chilled with liquid nitrogen. Mortars and pestles were cleaned before use with ethanol and baked at 200 °C for 2 h to destroy RNases. The laboratory bench and all other instruments used during grinding and RNA extraction were treated with RNaseZAP (Thermo Fischer Scientific, Waltham, Massachusetts, USA).

Ground tissue samples of seedling shoot tips, seedling root tips, leaves, roots and flowers were extracted using an RNeasy® Plant Mini kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. This extraction method relies on lysing the cells to release RNA, followed by a bind-wash-elute procedure using spin columns containing a silica membrane. RNA extraction from early and late pods was unsuccessful using the RNeasy® Plant Mini kit or TRIzol-reagent (Thermo Fischer Scientific, Waltham, Massachusetts, USA) using standard manufacturer protocols. Hence, RNA from these two tissues was extracted and purified using the Spectrum Plant Total RNA kit (Sigma-Aldrich, St Louis, Missouri, USA), following the manufacturer's protocol. This protocol is similar to the Qiagen RNeasy® Plant Mini kit in that it uses a spin column and a bind-wash-elute procedure, but uses different (proprietary) buffer solutions and also includes a DNase treatment before the RNA is eluted from the silica membrane. RNA concentrations of extracts from all tissue samples were estimated spectrophotometrically using a Nanodrop 1000 instrument (Thermo Fischer Scientific, Waltham, Massachusetts, USA).

4.2.2   Generation of TruSeq RNA libraries

Library generation, sequencing and transcriptome assembly were performed by The Genome Analysis Centre (TGAC), now the Earlham Institute (EI), Norwich, UK. Quality control of RNA samples was performed using Qubit dsDNA High Sensitivity and RNA High Sensitivity assays (Life technologies, Carlsbad, California, USA) and a 2100 Bioanalyzer instrument using the Agilent RNA 6000 kit. RNA libraries were constructed using the TruSeq RNA protocol (Illumina, San Diego, California, USA). In short: mRNA was extracted from 1 µg of purified RNA by poly-A pull-down-extraction using biotinylated beads. This RNA was fragmented and reverse transcribed into first-strand DNA, followed by second-strand synthesis. 3'-overhangs were removed through 3' to 5' exonuclease activity and blunt ends were formed by polymerase activity to fill in any 5' overhangs. A single 'A' nucleotide was added to the 3' ends of the double-stranded fragments, to create complementary overhangs with a single 'T' nucleotide on the 3'-end of indexing adapter sequences, allowing the DNA samples to be labelled with 6bp sample-specific adapter sequences before loading onto the flow-cell, allowing samples from different tissues to be pooled (see Table 9).

Table 9. Sequences of indexing adapters used to differentiate sequence originating from different tissues

Tissue of origin	Indexing adapter sequence
Seedling shoot tip	5' – CGATGT – 3'
Seedling root tip	5' – TGACCA – 3'
Leaf	5' – CAGATC – 3'
Root	5' – CCGTCC – 3'
Flower	5' – CTTGTA – 3'
Early pod	5' – GGCTAC – 3'
Late pod	5' – AGTCAA – 3'

Unligated adapters, adapter multimers and short fragments were removed by bead-based size selection using AMPure XP beads (Beckman Coulter, Brea, California, USA). To increase the concentration of DNA prior to sequencing and to enrich fragments that had been properly labelled on both ends, PCR was performed using the reverse complements of the relevant adapters as non-specific PCR-primers. Aliquots of the DNA libraries were run on a PerkinElmer GX instrument using the DNA High Sensitivity Reagent kit (PerkinElmer,

Waltham, Massachusetts, USA) to measure insert size. DNA concentration was measured using a Qubit dsDNA High Sensitivity assay.

4.2.3 Paired-end sequencing

DNA libraries were normalised by equal concentrations and pooled to a total concentration of 10 nM using 10 mM Tris-Cl buffer pH 8.5 (Qiagen, Hilden, Germany). This library pool was denatured using NaOH, by making a solution containing 2 nM DNA and 0.1 M NaOH. 5 μl of this solution were mixed with 995 μl HT1 buffer (Illumina, San Diego, California, USA). 120 μl of this diluted library pool were mixed with 1.2 μl PhiX Control v3 (Illumina, San Diego, California, USA) as a sequencing control. Clonal clustering of the flow-cell was performed by an Illumina cbot instrument using the TruSeq Paired-End Cluster Generation kit v3 and following the PE_amplification_Linearization_Blocking_PrimerHyb_v8 recipe (Illumina, San Diego, California, USA). Sequencing was performed using an Illumina HiSeq2500 system (Control Software 1.5.15.1 and RTA 1.13.48), following the manufacturer's instructions. One library pool containing DNA derived from the seedling shoot tip, seedling root tip, leaf, root and flower samples was run on duplicate flow-cell lanes using TruSeq SBS v3 High Sensitivity sequencing chemistry for 100 cycles of each paired-end read. A separate library pool containing seedling shoot tip, root, flower, early pod and late pod samples was on a single flow-cell lane using TruSeq SBS v4 High Sensitivity sequencing chemistry for 126 cycles of each paired-end read. Sequencing reads were de-multiplexed according to the 6 bp indexing adapters using the CASAVA 1.8 software package (Illumina, San Diego, California, USA) allowing for a single one-base-pair mismatch per library. Read files were converted into FASTQ format using the bcl2fastq conversion software (Illumina, San Diego, California, USA).

4.2.4 Quality control of sequencing data

Sequence quality control was performed using the FastQC algorithm (fastqc-0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to check for basic quality control metrics in the raw data including base quality scores, sequence length distribution and sequence duplication. The kmer-based contamination-screening pipeline Kontaminant (developed in-house by TGAC) (http://www.tgac.ac.uk/kontaminant/) was used to screen and filter for contamination in the raw reads.

The SortMeRNA algorithm (http://bioinfo.lifl.fr/RNA/sortmerna/) was used for filtering, mapping and operational taxonomic unit (OTU) picking of reads. SortMeRNA uses a file of

reads (fasta or fastq format) and one or several ribosomal RNA database file(s) as inputs and separates out rRNA and rejected reads from the mRNA reads.

Trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) was used to trim away adapter sequences. The commands used to run these algorithms are shown in Appendix 3.2.

4.2.5   Transcriptome assembly

RNA-Seq reads from the seven tissue/developmental stage-specific Illumina libraries were de novo assembled into separate transcriptomes using Trinity v2.1.1 (Grabherr et al., 2011). TransDecoder (http://transdecoder.github.io) was used to predict open reading frames and generate predicted proteins from within each set of assemblies. The software CD-HIT (Li and Godzik, 2006) was used to construct a non-redundant set of 95,793 protein sequences derived from the union of the seven sets. From these, 86,038 representative cognate transcript assemblies encoding these proteins were identified and used to construct a reduced reference sequence. Read counting was performed against this reference to measure transcript abundance across the seven tissues using the software package RSEM (Li and Dewey, 2011). Commands are shown in Appendix 3.3.

4.2.6   Automatic annotation of assembled transcripts

The results from the sequencing were processed using TGAC's in-house algorithm AnnotF, as used by De Vega et al. for the annotation of the red clover draft genome (De Vega et al., 2015). ORFs were predicted from each assembled contig, searching all six possible reading frames for triplets encoding methionine and stop codons. All potential ORFs comprising more than 100 codons were translated in-silico to generate a predicted proteome. The processed dataset was then screened using a BLAST2GO (Conesa et al., 2005) scan that compared the predicted proteins against the NCBI database in order to annotate each contig with its closest match by amino acid sequence homology. In addition, each predicted protein sequence was scanned using InterProSCAN 5 (Jones et al., 2014) to predict functionally characterised domains. Commands are shown in Appendix 3.4.

4.2.7   Reverse transcription of RNA from seven tissues of grass pea

The RNA samples that were extracted and purified for sequencing to produce the multi-tissue transcriptome were re-used to produce cDNA for cloning the ORFs of candidate genes. Besides the samples chosen for sequencing, all other RNA samples that had been

prepared in parallel and which also passed the quality check prior to sequencing (using the Qubit and Bioanalyzer systems as described) were reverse transcribed to generate cDNA.

The nucleic acid content of each sample was measured spectrophotometrically using a Nanodrop 8000 instrument (Thermo Fischer Scientific, Waltham, Massachusetts, USA). As these samples had already passed a quality control step no further DNAse treatment was performed on these samples.

First-strand cDNA-synthesis was performed using SuperScript II reverse transcriptase (Thermo Fisher Scientific). For each reaction 16 µl of template RNA, 2 µl of 10mM dNTPs (2.5mM of each nucleotide) and 2 µl 0.5 µg/µl oligo dT primers were mixed. The mix was incubated at 65 °C for 5 minutes, followed by 1 minute on ice. Separately, a master mix was prepared containing (for each sample): 4 µl 10X reverse transcriptase buffer, 8 µl 25 mM MgCl2, 2 ul 0.1 M DTT and 2 µl RNaseOUT. To each sample containing template RNA, dNTPs and primers, 18 µl of this master mix were added, followed by incubation at 42 °C for 2 minutes. Then, 2 µl of SuperScript II reverse transcriptase were added to each sample and the reactions were incubated at 42 °C for 50 minutes, followed by termination at 70 °C for 15 minutes. After chilling the reaction products on ice for 2 minutes, 2 µl of RNase H were added to each reaction followed by incubation at 37 °C for 20 minutes to degrade any RNA left over.

4.2.8   Design of primers for amplification of ODAP synthase candidate genes

The Gateway™ cloning system was used to clone the candidate genes. This cloning system relies on the specific and reversible exchange of att-stites between DNA sequences, catalysed by two clonase enzymes. Briefly, PCR products of the genes to be cloned are produced to include the attB1 and attB2 sites at their 5' and 3' ends, respectively. The attB sites (and the DNA fragment between them) recombine with the attP sites (and the DNA fragment between them) on the donor vector, catalysed by the BP clonase. The recombination of the attB and attP sites forms new sites called attL sites. These recombine with the attR sites on the destination vector, catalysed by the LR clonase, which again forms attB sites on the expression vector including the gene to be expressed.

All forward primers were designed to include the following sequence at their 5' ends:

5' – GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG – 3'

This sequence includes the attB1 binding site for Gateway™ recombination, a Shine-Dalgarno sequence to allow the genes to be expressed in bacteria (not performed as part of

this study), a Kozak sequence for expression in eukaryotes and a start codon. This sequence was followed by 18-23 gene-specific nucleotides downstream of the start codons of the candidate genes which served as specific PCR-primer regions. In the case of three of the candidate genes (BAHD 9, BAHD 10 and BAHD 11) the sequence immediately following the start codon was a poor region for PCR-priming with low (< 33 %) GC content. In order to allow efficient amplification of these sequences, the 18 bases following the start codon, were substituted with alternative codons that are common in *Pisum sativum*. This sequence was added in between the upstream sequence required for cloning and the gene specific priming region following this sequence. The inserted 18 nucleotides were the same for all three candidate genes where this issue occurred.

sequence following start codon (native):   GAT TCC GTG AAA GTA ATA

in amino acids: D   S   V   K   V   I

alternative codons (used in the primer):   GAC TCA GTC AAG GTT ATC

All reverse primers were designed to include the following sequence at their 5' ends:

5' – GGGGACCACTTTGTACAAGAAAGCTGGGTTTTA – 3'

This sequence includes the reverse complements of the attB2 binding site and a stop codon. This stop codon is not strictly necessary, as the amplified region also included the native stop codon of the candidate genes. Gene specific primer regions were selected from the ends of the coding sequence for each candidate gene, up to and including the stop codon. In cases where the end of a candidate gene did not form an acceptable primer region, a primer region up to 100bp downstream of the stop codon was selected.

Gene specific regions were selected to have predicted melting temperatures between 54 °C and 57 °C. All complete oligonucleotides were checked for self-annealing and secondary structures using the Integrated DNA technologies OligoAnalyzer 3.1 tool (https://www.idtdna.com/calc/analyzer). The sequences of the final oligonucleotides used are shown in Table 10.

Table 10. Oligonucleotides designed for the amplification of coding sequences of candidates for genes encoding ODAP-synthases from grass pea cDNA. Gene models selected for inclusion in the transient expression experiment are marked with *. The forward primer regions for candidates BAHD10 and BAHD11 were identical, as were the reverse primer regions for candidates BAHD5, BAHD10 and BAHD11. Because of this, only one primer pair was used to attempt to amplify both BAHD10 and BAHD11.

primer	target gene model	attB1 sequence	codon exchange spacer	gene specific region
BAHD1-F	Shoot_1_m.69767	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		GCAAACATTCGAGTCATCT
BAHD2-F	Flower_1_m.81855	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		AGTTCAGTCCAAGTTCTTTC
BAHD3-F	Early_pod_2_m.99813 *	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		AGTTCCATCCAAATCCTCTC
BAHD4-F	Leaf_1_rep_2_m.77806	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		TCTTCCATCAAACTCATATCAAC
BAHD5-F	Radicle_3_m.76038	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		CCTCACCTTTCCACACTCA
BAHD6-F	Radicle_3_m.56696	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		GATACCGTTCGAGTCATCT
BAHD7-F	Late_pod_1_m.7957	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		GATGGTTCAGTGAGAGTAAT
BAHD8-F	Root_3_113511 *	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG		GGTTCCGTGAAAGTAATATCC
BAHD9-F	Early_pod_1_m.70389 *	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG	GACTCAGTCAAGGTTATC	TCCACAACAACAATCCAAGCA
BAHD10/11-F	Early_pod_1_m.70388 *	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG	GACTCAGTCAAGGTTATC	TCCACAACAACAATCCAAG
BAHD10/11-F	Root_3_m.101418 *	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATG	GACTCAGTCAAGGTTATC	TCCAACAACAATCCAAG

primer	target gene model	attB2 sequence	gene specific region
BAHD1-R	Shoot_1_m.69767	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	TGGAGTAATGGCATGCATC
BAHD2-R	Flower_1_m.81855	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	ACAAATGAAAAGCTGTGAAAA
BAHD3-R	Early_pod_2_m.99813 *	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	ACCAGAAGCAGCATCCATA
BAHD4-R	Leaf_1_rep_2_m.77806	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	ACAATACATAACCCTTCATAACA
BAHD5.10.11-R	Radicle_3_m.76038	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	CTCTCTCTCAAACACACAT
BAHD6-R	Radicle_3_m.56696	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	CCAACCAACCGTATTCAAAA
BAHD7-R	Late_pod_1_m.7957	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	AACAATGATAGGTTCTTCCC
BAHD8-R	Root_3_113511 *	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	ACCAGACACAACCAAGCA
BAHD9-R	Early_pod_1_m.70389 *	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	CAACCAAGCACAAAACACA
BAHD5.10.11-R	Early_pod_1_m.70388 *	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	CTCTCTCTCAAACACACAT
BAHD5.10.11-R	Root_3_m.101418 *	GGGGACCACTTTGTACAAGAAAGCTGGGTTTA	CTCTCTCTCAAACACACAT

4.2.9   PCR amplification of candidate BAHD-acyltransferases from cDNA

PCR reactions were set up to amplify the candidate ORFs for cloning into expression vectors. For each candidate, cDNA derived from the grass pea tissue with the highest expression level for this candidate (according to RSEM read counts) was used (see Table 11).

Table 11. Tissue origin of cDNA used for amplification of ODAP-synthase candidates

Gene model	ID	cDNA source
Shoot_1_m.69767	BAHD 1	seedling root tip
Flower_1_m.81855	BAHD 2	Seedling shoot tip
Early_pod_2_m.99813	BAHD 3	Late pod
Leaf_1_rep_2_m.77806	BAHD 4	Late pod
Radicle_3_m.76038	BAHD 5	Seedling root tip
Radicle_3_56696	BAHD 6	Root
Late_pod_1_m.7957	BAHD 7	Late pod
Root_3_m.113511	BAHD 8	Seedling shoot tip
Early_pod_2_m.70389	BAHD 9	Seedling shoot tip
Early_pod_2_m.70388	BAHD 10	Seedling shoot tip
Root_3_m.101418	BAHD 11	Seedling shoot tip

For each PCR reaction, 1 µl of template cDNA was mixed with 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM), 2 µl DEPC-treated water and 5 µl of 2X GOTaq® G2 Green master mix containing polymerase, proprietary G2 reaction buffer, 1.6 mM total dNTPs and yellow and blue loading dyes (Promega, Madison, Wisconsin, USA). PCR reactions were run according to the protocol shown in Table 12.

Table 12. PCR protocol for the amplification of candidate ODAP-synthases from grass pea cDNA

Step		Temperature	Duration
First denaturation		95 °C	2 min
40 cycles	denaturation	95 °C	30 sec
	annealing	52 °C	30 sec
	elongation	72 °C	1 min 30 sec
Final elongation		72 °C	10 min

PCR products were purified using a QIAquick PCR purification kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. Briefly, this purification method uses a bind-wash-elute procedure in spin-columns containing a silica membrane to remove primers, nucleotides, enzymes and salts from the PCR product. This was followed by PEG 8000 precipitation to remove primer dimers and any other smaller DNA fragments from the PCR products. A PEG 8000 solution containing 30 mM $MgCl_2$ was prepared and sterilised by filtration through a 0.45 µm sterile filter. To precipitate DNA fragments >400 bp, 60 µl of this solution were mixed with 90 µl of TE buffer (pH 8.0) and 30 µl PCR product and the mixture was centrifuged in a benchtop centrifuge at 16,250 g for 20 minutes. The supernatant was discarded and the pellet resuspended in 30 µl 10 mM Tris-Cl buffer pH 8.5 (Qiagen, Hilden, Germany). To further purify PCR products of the right size, 20 µl of this solution were loaded onto a 1 % w/v agarose gel, which was set to run for 1 h at 35 V to separate DNA fragments by size. Bands at the expected size of ~1300 bp were excised from the gel under brief UV-illumination. DNA was extracted from these gel bands using a QIAquick gel extraction kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. Briefly, the agarose slice was solubilised, followed by a bind-wash-elute procedure using spin-columns containing silica membranes to remove agarose, ethidium bromide and salts. The concentration of purified fragments was estimated spectrophotometrically using a Nanodrop 8000 instrument (Thermo Fischer Scientific, Waltham, Massachusetts, USA).

4.2.10 Amplification of destination vectors

In order to culture cells harboring the destination vectors and the donor vector, it was necessary to use a strain of *Escherichia coli* (such as DB3.1) resistant to the ccdB gene contained within the cloning site of the vector, as this gene is lethal to most strains of *E. coli*, allowing for the selection of successful constructs at a later stage. Competent DB3.1 cell suspension (20 µl), 1 µl of pEAQ-HT-DEST1 plasmid carrying either the Gateway™ cloning attachment sites or a GFP gene as a positive control and 19 µl of dH2O were mixed and incubated on ice for 30 minutes. The suspension was then subjected to a heat shock by submersion in a 42 °C water bath for 90 seconds followed by incubation at 37 °C for 1 hour. Of the suspension, 20 µl and 100 µl were then plated onto lysogeny broth (LB) agar plates containing 10 µg/ml gentamycin or 50 µg/ml kanamycin for the pDONR207 donor vector or the pEAQ-HT-DEST1 destination vector respectively. Plates were incubated overnight at 37 °C. Single colonies from these plates were inoculated into 10 ml liquid LB medium containing the same concentration of antibiotic. These liquid cultures were again incubated

overnight at 37 °C in a shaking incubator. Plasmids were extracted from these cultures using a QIAquick spin miniprep kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Briefly, bacteria are lysed and lysates cleared by centrifugation, followed by a bind-wash-elute procedure using spin-columns containing a silica membrane. Separate liquid cultures (2 ml) were mixed with equal volumes of 40 % glycerol for long-term storage at -80 °C. Glycerol stocks of the entry clones were kept for future use and as a back-up.

4.2.11 Assembly of expression clones using Gateway™ recombination

To generate entry clones containing the cDNA of the candidate genes, 7 µl of PCR product in 10 mM Tris-HCl buffer pH 8.5 (Qiagen, Hilden, Germany) were mixed with 1 µl pDONR207 (see map in Figure 58) in 10 mM Tris-HCl buffer pH 8.5 (150 ng/µl). To this mixture, 2 µl of BP Clonase II enzyme (Sigma-Aldrich, St Louis, Missouri, USA) were added and the reaction was incubated at room temperature overnight. The reaction was stopped by adding 1 µl proteinase K solution and incubating for 10 minutes at 37 °C. The entry clones were transformed into *E. coli* DH5α (a ccdB-sensitive strain) by mixing 2 µl of the reaction product with 50 µl of competent cell suspension, incubating the cells on ice for 60 minutes, followed by a heat shock in a water bath at 42 °C for 90 seconds. The use of a ccdB-sensitive strain at this point ensures that any bacteria transformed with the unchanged destination vector (not including the cloned gene, but still containing the ccdB gene) are rendered unviable. After the heat shock, 250 µl of SOC medium (2 % w/v bacto-tryptone, 0.5% w/v bacto yeast extract, 0.05 % w/v NaCl, 2.5mM KCl, 0.01M $MgCl_2$, 20mM glucose) were added to each transformation and the cells were allowed to recover in a shaking incubator at 37 °C for 1 h. Of the suspension, 20 µl and 100 µl were then plated onto LB agar plates containing 10 µg/ml gentamycin. These plates were incubated at 37 °C overnight.
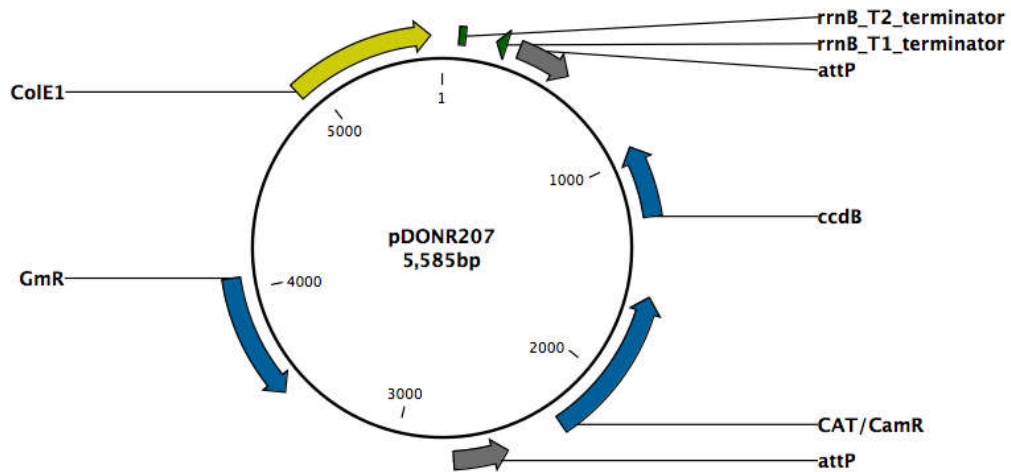
Figure 58. Map of the pDONR207 donor vector. Annotated features (clockwise): rrnB_T1, rrnB_T2 – transcriptional terminators; attP – Gateway cloning sites; ccdB – selection gene lethal to F⁻ strains of E.coli; CAT/CamR – chloramphenicol acetyltransferase (chloramphenicol selectable marker); GmR – gentamycin resistance gene (selectable marker); ColE1 - bacterial replication origin. Drawn using CLCbio Main Workbench 6.5 (https://www.qiagenbioinformatics.com/)

Bacterial colonies were picked from each plate using a toothpick and transferred into 100ul of dH$_2$O. These suspensions were incubated at 95 °C for 10-15 minutes to extract DNA for PCR. PCR reactions were set up by mixing 3 μl of extracts with 1 μl attL1 primer, 1 μl attL2 primer (both 10 μM) and 5 μl GOTaq G2 Green Master Mix (Promega, Madison, Wisconsin, USA).

attL1 primer: 5' – TCGCGTTAACGCTAGCATGGATCTC – 3'

attL2 primer: 5' – GTAACATCAGAGATTTTGAGACAC – 3'

Fragments were amplified using the PCR protocol shown in Table 13. The size of PCR products was checked by running 4 μl of each PCR product on a 1 % w/v agarose electrophoresis gel containing 0.01 % v/v ethidium bromide.

Table 13. Colony PCR protocols to amplify inserts from entry clones

Step		Temperature	Duration
First denaturation		95 °C	2 min
40 cycles	denaturation	95 °C	30 sec
	annealing	52 °C	30 sec
	elongation	72 °C	1 min 30 sec
Final elongation		72 °C	10 min

The remaining 6 µl of the sample with the expected product size was purified using a QIAquick PCR purification kit (Qiagen, Hilden, Germany), as described in section 4.2.9. The purified product was eluted in 34 µl of 10 mM Tris-Cl pH 8.5 buffer. To confirm the identity of the cloned sequences, 17 µl of each purified PCR product were mixed with 2 µl of attL1 primer (10 µM), while the remaining 17 µl were mixed with 2 µl attL2 primer. These mixtures were sent for sequencing (Eurofins, Luxembourg).

The forward and reverse sequences obtained were combined into single contigs using the CLCbio Main Workbench package 6.5 (https://www.qiagenbioinformatics.com/), using sequencing traces to resolve conflicts between the forward and reverse sequences. The assembled contigs were compared to the cDNA sequences retrieved from the transcriptome to check for amplification errors.

The liquid cultures derived from the single colonies that gave rise to the highest quality sequences were selected for further processing. Plasmids were extracted from these cultures using a plasmid miniprep kit (Qiagen, Hilden Germany) as described in section 4.2.10. Due to the quantity of liquid culture (10 ml), double volumes of buffers were used to suspend and lyse the cells. The lysate was applied to the spin columns in two batches, discarding the liquid between centrifugations. The concentration of extracted entry clone plasmids was estimated by spectrophotometric measurement using a Nanodrop 8000 instrument (Thermo Fischer Scientific, Waltham, Massachusetts, USA).

Expression clones were assembled using the LR reaction to combine the coding sequences carried on the entry clones with the regulatory sequences on the pEAQ-HT-DEST1 destination vector (see map in Figure 59). For each reaction, 150 ng of entry clone were mixed with 250 ng of destination vector and TE buffer (pH 8.0) to a volume of 8 µl. On ice, 2 µl of Gateway LR Clonase II mix (Invitrogen) were added to these solutions. The reactions were incubated for 25 °C for 1 h. Reactions were stopped by adding 1 µl proteinase K and

incubating at 37 °C for 10 minutes. *E. coli* DH5α cells were transformed with the reaction product as described for the products of the BP reaction. After allowing the transformed cells to recover in SOC medium, 20 µl and 100 µl of the cell suspension were spread onto LB agar plates containing 50 µg/ml kanamycin. The plates were incubated at 37 °C overnight.
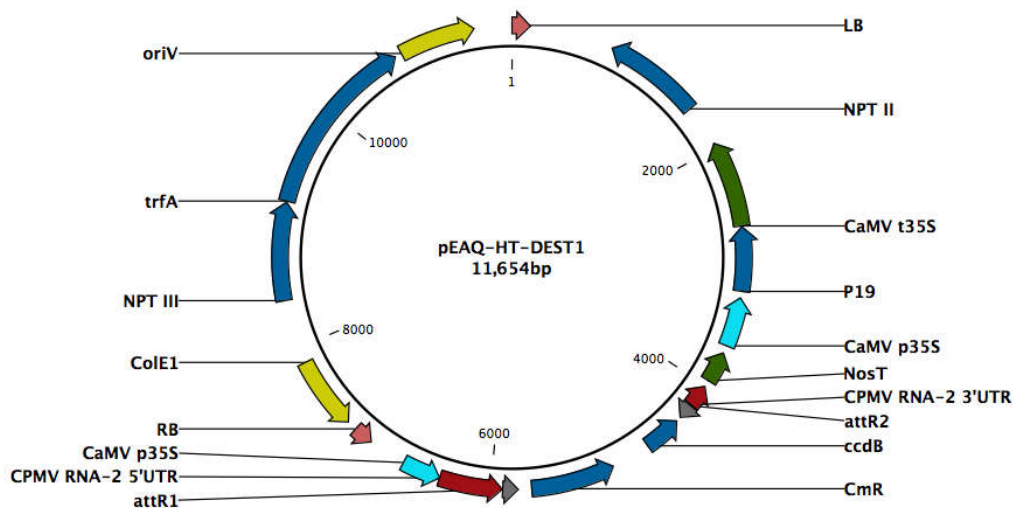
Figure 59. Map of the pEAQ-HT-DEST1 destination vector. Annotated features in clockwise order: LB, RB – left and right border of the T-DNA; NPT II, NPT III – neomycin phosphotransferase (kanamycin selectable marker); CaMV t35S – Cauliflower Mosaic Virus 35S transcriptional terminator; p19 – inhibitor of gene silencing; CaMV p35S Cauliflower Mosaic Virus 35S promoter; NosT – nopaline synthase transcriptional terminator; CPMV RNA-2 3'UTR, CPMV RNA-2 5'UTR – Cowpea Mosaic Virus RNA-2 untranslated regions (modified); attR2, attR1 – Gateway cloning sites; ccdB – selection gene lethal to F⁻ strains of E.coli; CmR – chloramphenicol resistance gene (selectable marker); ColE1 - bacterial replication origin; trfA – plasmid RK2 replication initiation protein; oriV – eukaryotic replication origin. Drawn using CLCbio Main Workbench (https://www.qiagenbioinformatics.com/)

Single colonies were picked from these plates and used to inoculate 10 ml LB liquid cultures containing 50 µg/ml kanamycin, which were incubated at 37 °C overnight. In addition, cells from the same colonies were suspended in 100 µl water and incubated at 95 °C for 10 minutes. The incubated suspension was used as a template for colony PCR, by mixing 3 µl of it with 1 µl attB1 primer (10 µM), 1 µl attB2 primer (10 µM) and 5 µl GOTaq G2 Green Master Mix. PCR reactions were run using the protocol shown in Table 14. The identity of PCR products was confirmed by size using gel electrophoresis and sequencing from both ends using the attB primer pair, as described above.

Table 14. Colony PCR protocols to amplify inserts from expression clones

Step		Temperature	Duration
First denaturation		95 °C	2 min
33 cycles	denaturation	95 °C	30 sec
	annealing	50 °C	30 sec
	elongation	72 °C	1 min 30 sec
Final elongation		72 °C	10 min

Expression clones were extracted from overnight cultures using a QIAquick plasmid miniprep kit (Qiagen, Hilden, Germany), as described in section 4.2.10.

4.2.12 Transformation of *Agrobacterium tumefaciens* by electroporation

Electrocompetent *Agrobacterium tumefaciens* cells of the strain GV3101 pMP90 (Koncz and Schell, 1986) were taken from storage at -80 °C and allowed to thaw on ice. 50 μl of bacterial suspension were gently mixed with 1 μl of plasmid solution (~150 ng DNA/μl). As a positive control of transient expression in *N. benthamiana*, an additional sample of 50 μl of electrocompetent GV3101 cells was mixed with 4 μl of pEAQ-HT-GFP solution (15 ng DNA/μl). The suspensions were transferred into electroporation cuvettes with a gap size of 2 mm. Electroporation was performed using a GenePulser device (Bio-Rad, Hercules, California, USA) set to a voltage of 2.5 kV with capacitance of 25 μF and resistance of 400 Ω. Time constants were recorded as 9.9 msec or 10.0 msec. Immediately following electroporation, 1 ml of L medium was added to each cuvette and the cells were allowed to recover on ice for 30 minutes followed by incubation in a shaking incubator at 28 °C for 2 hours. Aliquots of 10 μl and 100 μl of the transformed bacterial suspension were spread on LB-agar plates containing 10 μg/ml rifampicin and 50 μg/ml kanamycin. These plates were incubated for two days at 30 °C. Single colonies were picked from these plates and inoculated into 10 ml LB medium containing 50 μg/ml kanamycin and 10 μg/ml rifampicin. The cultures were incubated for three days in a shaking incubator at 28 °C. From these liquid cultures, glycerol stocks were produced by mixing 1 ml of liquid culture with 1 ml of 40% glycerol solution. Glycerol stocks were stored at -80 °C. Of the remaining liquid cultures, 10 μl were transferred into fresh LB medium containing 50 μg/ml kanamycin and 10 μg/ml gentamycin and the new cultures were incubated at 28 °C in a shaking incubator overnight. To confirm that the cells in these cultures carried the candidate genes, PCRs using a primer pair targeting the attB sites of the expression plasmids were performed. To

release the plasmids 1µl of liquid culture was mixed with 30 µl of RO-water and incubated at 95 °C for 15 minutes. An aliquot (3 µl) of this solution was used as a template in a PCR reaction analogous to the one described in Table 14. Reaction products were run on a 1% w/v agarose electrophoresis gel.

4.2.13   Agroinfiltration of candidate gene expression vectors

The Agrobacterium cultures grown overnight were spun down in a tabletop centrifuge at 2235 g for 15 minutes to pellet the cells. After removing the supernatant, the cells were resuspended in agroinfiltration solution (10 mM $MgCl_2$, 10 mM 2-(N-morpholino)ethanesulfonic acid (MES) and 200 µM acetosyringone in distilled water, adjusted to pH 5.6 using potassium hydroxide). Acetosyringone increases transformation efficiency by inducing the *Agrobacterium vir* gene cluster, which is involved in mediating gene transfer into the plant cells (Godwin et al., 1991). The suspensions were incubated at room temperature on a tilting table for 3 hours then diluted in Agroinfiltration solution to OD_{600} = 0.2 prior to infiltration.

Nicotiana benthamiana plants were grown up in a glasshouse during June (average 16 hours of daylight/day) for 3-4 weeks in well-watered conditions. For each infiltration, a small nick was made on the abaxial side of a leaf, using a sterile needle, taking care not to puncture the entire leaf, but only to open up the mesophyll intercellular space (Figure 60A). A syringe containing the bacterial suspension was then placed onto this nick and the suspension was gently pushed into the leaf mesophyll intercellular space until at least 90% of the leaf area was infiltrated (Figure 60B). In some cases, up to three injection sites were needed to achieve this. Each leaf was infiltrated with a total of 400 - 800 µl of bacterial suspension, three leaves of each plant were infiltrated and each leaf was treated as one biological replicate. Each plant was injected with only one suspension to avoid cross-contamination.
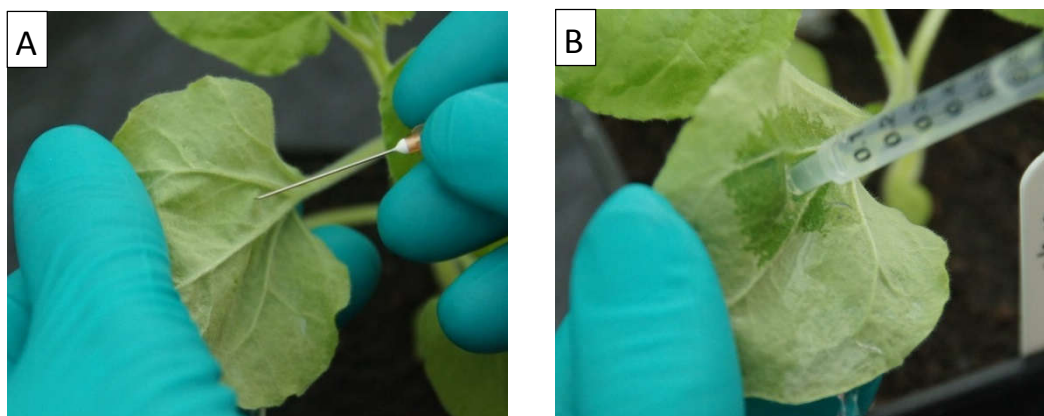
Figure 60. Photographs of Agrobacterium infiltration into Nicotiana benthamiana leaves. A) a small nick is made on the abaxial side of the leaf. B) The Agrobacterium suspension is pushed into this nick and percolates through the mesophyll intercellular space

Three days after infiltration with *Agrobacterium* suspension, the same leaves were infiltrated with metabolite solutions, using the same methodology. Plants were infiltrated with either 1mM L-DAP in water (pH 6 adjusted with KOH); 1mM L-DAP and 1mM oxalic acid in water (pH 6, adjusted with KOH) or water (mock). Each plant was injected with only one metabolite solution to avoid cross-contamination.

Entire infiltrated leaves were harvested five days after agroinfiltration, i.e. two days after infiltration with metabolite solutions and flash-frozen in liquid nitrogen. Leaf samples were ground using a mortar and pestle, chilled with liquid nitrogen, followed by 24 h freeze drying (BenchTop SLC®, Virtis, Gardiner, New York, USA). 5 mg of each sample were weighed out. Free amino acids were extracted by adding 500 µl of 70% ethanol in water (HPLC-grade) and shaking the suspension overnight at 20°C. The suspensions were centrifuged for 30 min at 16,250 g in a benchtop centrifuge. The supernatant was transferred into a fresh microcentrifuge tube and the pellet was resuspended in another 500 µl 70% ethanol and left to extract for another 30 minutes before centrifugation. The supernatant was again transferred and the extraction was repeated a third time. The supernatants of all three extraction steps were combined and evaporated to dryness in a GeneVac (EZ-2 Elite, Genevac, Ipswich, UK). The dried extracts were re-dissolved in 1 ml of RO-water by vortexing and incubation at 55 °C for 20 minutes.

Standard solutions of β-L-ODAP and L-DAP.HCl containing 50 mg/ml of the respective compounds were prepared and processed in parallel with the sample extracts using the AccQ-Tag derivatisation procedure (Waters, Milford, Massachusetts, USA), as in section 3.2.9. Dissolved sample extract (20 µl) or standard solution were mixed with 60 µl of AccQ-Tag borate buffer. To this, 20 µl of dissolved AccQ-Tag reagent were added, mixed

immediately and incubated at 55 °C for 10 minutes. The derivatised samples were diluted 1:1500 before LCMS injection.

4.2.14 LCMS measurement of L-DAP and β-L-ODAP in derivatised *N. benthamiana* extracts

L-DAP and β-L-ODAP in samples and standards were measured using a Xevo triple quadrupole TQ-S instrument (Waters, Milford, Massachusetts, USA). A volume of 5 µl of each sample was injected into a Kinetex 2.6 µm EVO C18 100 Å 100 x 2.1 mm column with a C18 guard column (Phenomenex, Macclesfield, UK). The solvent profiles for this experiment are shown in Table 15.

Table 15. Solvent profiles used for the injection of derivatised N. benthamiana extracts. Linear gradients were used for all changes in solvent mixtures.

Time (min)	Flow rate (ml/min)	Water + 0.1% formic acid (HPLC-grade)	Acetonitrile (HPLC-grade)
0	0.6	99%	1%
0.4	0.4	99%	1%
7.0	0.4	75%	25%
8.5	0.4	10%	90%
9	0.4	10%	90%
9.1	0.4	99%	1%
12.6	0.4	99%	1%

Four mass transitions (347.1 u → 116.1 u; 347.1 u → 145.1 u; 347.1 u → 171.1 u and 347.1 u → 303.1 u) were used to identify the presence of ODAP. These mass transitions were the main fragmentations observed by the Xevo TQ-S automatic calibration protocol, as described in the previous chapter 3.2.9. The transition 347.1 u → 171.1 u which represents the singly charged AccQ-Tag-derivatised molecule fragmenting to release the derivatisation group is the most prevalent transition. Both β-L-ODAP and α-ODAP are characterised by the same mass transitions, but the differences in structure cause the two derivatised compounds to elute at different retention times on the HPLC column, allowing them to be distinguished.

4.3 Results and discussion

4.3.1 Quality of extracted RNA

The RNA extracted from grass pea tissues using the Qiagen Total RNA kit (Thermo Fischer Scientific, Waltham, Massachusetts, USA) was tested using the Qubit dsDNA High Sensitivity and RNA High Sensitivity assays and the 2100 Bioanalyzer instrument (Life technologies, Carlsbad, California, USA) as detailed in section 4.2.2. RNA concentrations and quality are shown in Table 16. One sample from each seedling shoot tip, seedling root tip, whole root and flower tissues was selected for sequencing. All leaf, early pod and late pod samples failed the quality control criteria.

*Table 16. Concentration and quality of RNA samples. * samples selected for sequencing, ** this sample was initially rejected, but selected for sequencing after review of the quality control results*

Sample ID	RNA conc. (ng/ul)	DNA conc. (ng/ul)	total RNA (ng)	RNA quality score	Pass/Fail
Shoot 1 *	**1259.3**	**70.487**	**50372**	**10**	**Pass**
Shoot 2	1205.3	69.295	48212	10	Pass
Shoot 3	810.68	66.25	32427.2	10	Pass
Radicle 1	455.58	56.252	18223.2	10	Fail
Radicle 2	664.21	58.575	26568.4	10	Pass
Radicle 3 *	**708.75**	**59.465**	**28350**	**10**	**Pass**
Leaves 1 * *	**49.839**	**6.7391**	**1993.56**	**8.3**	**Fail**
Leaves 2	7.8587	0.73252	314.348	NA	Fail
Leaves 3	0.7261	0.53912	29.044	5.6	Fail
Roots 1	173.95	61.559	6958	9.4	Fail
Roots 2	182.15	70.336	7286	9.6	Fail
Roots 3 *	**186.81**	**71.629**	**7472.4**	**9.4**	**Pass**
Flowers 1 *	**1652.6**	**70.99**	**66104**	**10**	**Pass**
Flowers 2	1434.9	72.355	57396	10	Pass
Flowers 3	1552.9	73.086	62116	10	Pass
Early Pod 1	18.085	0.51213	723.4	7.5	Fail
Early Pod 2	1.0065	0.79548	40.26	6.5	Fail
Early Pod 3	22.157	0.92816	886.28	7.2	Fail

Late pod 1	18.995	1.3285	759.8	1.7	Fail
Late pod 2	22.616	1.8884	904.64	2.4	Fail
Late pod 3	14.087	0.97089	563.48	NA	Fail

To replace the failed samples I extracted additional samples of leaves, early pods and late pods using the Qiagen RNA extraction kit. The pod samples showed very low concentrations of nucleic acids after DNase clean-up when measured using a Nanodrop 1000 instrument (not shown). Hence, only the leaf samples were analysed by the RNA quality control pipeline.

Table 17. Concentration and quality of leaf RNA samples prepared to replace failed samples, neither sample fulfilled the selection criteria

Sample ID	RNA conc. (ng/µl)	DNA conc. (ng/µl)	total RNA (ng)	RNA quality score	Pass/Fail
Leaves 4	231.36	4.8598	9254.4	6.6	Fail
Leaves 5	278.83	6.3381	11153	5.2	Fail

Neither sample passed the quality control criteria. However, despite narrowly failing the initial quality control step, the original leaf 1 sample (see Table 16) was judged to be of sufficient quality for sequencing. To replace the failed pod samples, I used the Spectrum Plant total RNA kit (Sigma-Aldrich, St Louis, Missouri, USA) to extract RNA from these tissues. These extractions produced RNA of sufficient quality for sequencing as shown in Table 18.

*Table 18. Concentration and quality of early pod and late pod RNA samples prepared using the Spectrum RNA kit to replace failed samples. * samples selected for sequencing*

Sample ID	RNA conc. (ng/µl)	DNA conc. (ng/µl)	total RNA (ng)	RNA quality score	Pass/Fail
Early pod 5 *	266.69	59.15	13335	7.6	Pass
Late pod 4 *	255.41	46.50	12771	8	Pass

4.3.2 Summary of sequencing results

RNA samples were reverse transcribed into cDNA as described in section 4.2.2. The cDNA samples were sequenced using Illumina paired-end sequencing. Raw results of the sequencing are shown in Table 19. Seedling shoot tip, root and flower samples were run as part of three pooled sequencing lanes, while seedling root tip and leaf samples were run as part of two.

Table 19. Summary of results of the Illumina sequencing of cDNA derived from seven grass pea tissue samples.

Tissue	Number of reads	Read 1 mean Q30 to base	Read 2 mean Q30 to base	Mean insert size less adaptors
Seedling shoot tip	60,668,950	94	94	277
	63,109,728	94	89	
	47,452,003	126	124	
Seedling root tip	63,059,659	99	99	287
	65,837,884	94	99	
Leaf	40,388,718	94	94	262
	42,133,522	94	94	
Root	25,140,966	94	99	228
	26,096,845	94	94	
	42,615,758	126	126	
Flower	16,462,704	99	99	209
	17,266,885	94	99	
	45,928,837	126	126	
Early pod	48,709,681	126	126	223
Late pod	44,779,290	126	126	231
Total	**649,651,430**	**106**	**106**	**246**

4.3.3 Transcriptome assembly

The combined data obtained from each tissue were used for de novo assembly. Summary results of the assembly, ORF-prediction and automatic annotation are shown in Table 20. Tissue-specific transcriptomes and a non-redundant transcriptome based on all tissues were uploaded to a server that allowed sequence retrieval and BLAST-searches. Predicted proteomes based on the ORF-prediction were also uploaded as tissue-specific and non-redundant databases.

Table 20. Summary of results of the Trinity de novo assembly of RNAseq reads. The transcriptome of each tissue was assembled separately. Depth of coverage is given as average fold coverage of each assembly counting only aligned reads.

tissue	Assem-bled contigs	Longest contig	Mean contig length	N50	Total Bases in contigs	Depth of coverage	Coding regions	Annotated regions
Seedling shoot tip	369,248	15,584	602	1,439	222,323,426	72.3	118,622	74,163
Seedling root tip	312,081	15,618	653	1,458	203,973,440	60.4	115,098	60,790
Leaf	285,517	14,605	697	1,508	199,227,892	39.0	122,139	66,804
Root	347,994	15,587	637	1,380	221,941,192	39.3	127,484	67,882
Flower	329,169	13,599	588	1,313	193,834,760	36.4	114,207	70,945
Early pod	247,171	15,583	698	1,432	172,617,501	27.0	108,868	69,160
Late pod	238,446	15,579	710	1,440	169,512,820	25.2	107,490	68,461

These databases allowed me to search for grass pea transcripts and predicted proteins by homology and retrieve sequence data based on automatic annotations. I used these tools to identify candidates for genes coding for enzymes involved in the biosynthesis of β-L-ODAP.

4.3.4 Identification of a candidate gene encoding an oxalyl-CoA synthetase in grass pea

Performing protein BLAST searches (blastp) (Altschul et al., 1997) using both amino acid sequences of the *Arabidopsis* and the *Medicago* AAE3 enzymes against the non-redundant predicted protein database from grass pea resulted in a single significant hit, Early_pod_2_m.95177. This predicted protein shares 75% amino acid identity with AtAAE3 and 88% identity with MtAAE3 (alignment shown in Figure 61).
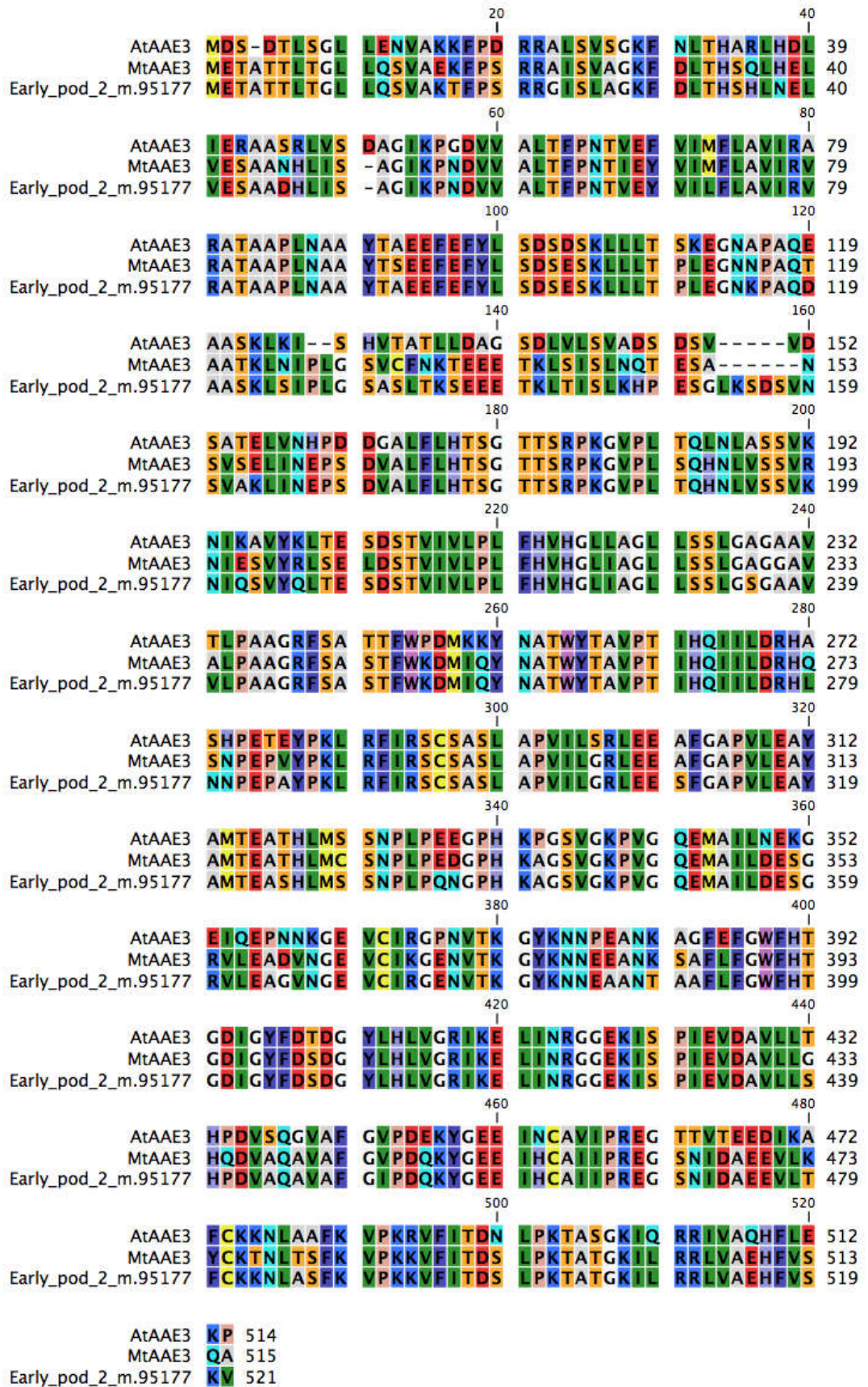
Figure 61. Alignment of amino acid sequences of oxalyl-CoA sythetases from Arabidopsis thaliana and Medicago truncatula and their closest homolog in grass pea. Residues are coloured according to RasMol colour scheme. Alignment produced and plotted using CLCbio Main Workbench 6.5 (https://www.qiagenbioinformatics.com/)

Due to the high degree of conservation of this pathway across eukaryote species and the close homology between the predicted protein from grass pea and the known oxalyl-CoA synthetases from *Arabidopsis* and *Medicago*, this is a promising candidate for the oxalyl-CoA synthetase in grass pea (Malathi et al., 1970). The corresponding transcript is expressed across all the tissues that I measured, with expression being highest in the flowers, early and late pods and lowest in leaves (see Figure 62).
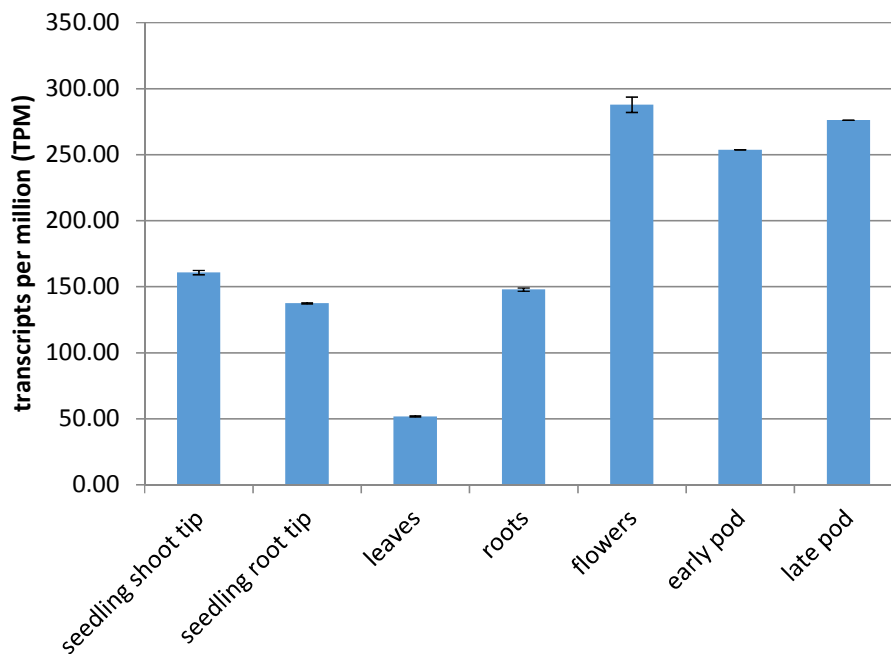
Figure 62. Transcript abundance of the putative Lathyrus sativus oxalyl-CoA synthetase by tissue, based on read-counting using the RSEM algorithm

If the identity of this gene as the oxalyl-CoA synthetase in grass pea can be confirmed experimentally, this would be a major step towards understanding the genetic basis of β-L-ODAP production.

4.3.5   Identification of transcripts encoding putative BAHD-acyltransferases

I extracted the sequences of BAHD acyltransferases from the non-redundant set of predicted proteins of grass pea by searching the automatic annotation results produced by the AnnotF algorithm for general terms such as "bahd acyltransferase dcr-like", or specific terms such as "vinorine synthase" or "hydroxycinnamoyl transferase". However as each contig was only annotated with one homology search result, rather than a tree of search results containing generic as well as more specific terms, I could not be sure of capturing all the contigs that had been annotated as BAHD acyltransferases without a complete list of possible specific search terms that would fall within this category. I therefore decided to

instead search the Interpro domain annotations for domains that are common to BAHD acyltransferases (see Figure 63). To this end, I retrieved all entries annotated as containing the chloramphenicol acetyltransferase-like domain (IPR023213) and manually curated this list to remove entries with homology to enzymes outside of the BAHD-AT superfamily. Genes of unknown homology were retained.

There are two conserved motifs common to BAHD-ATs. The first is the catalytically active HXXXD motif (typically followed by a glycine) which occurs near the middle of the amino acid sequence. The second is the less conserved DFGWG motif near the C-terminus of most BAHD acyltransferases. The function of this motif is unclear as it is not physically located near the catalytically active site in BAHD-ATs with known crystal structures, but it seems to have an important function in maintaining the structure of the enzyme (Ma et al., 2005).
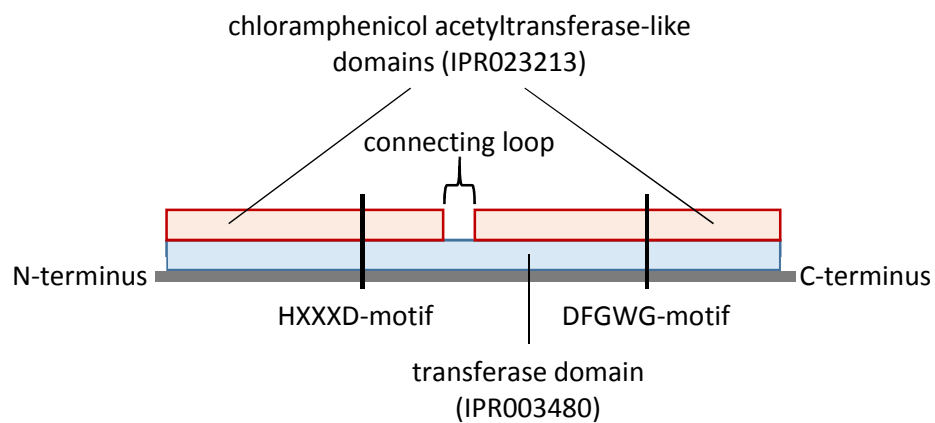
Figure 63. Generic domain architecture commonly found in BAHD-acyltransferases. Two domains, each with structural similarity to chloramphenicol acetyltransferase, are connected by a ~13 amino acid loop (Ma et al., 2005). The HXXXD and DFGWG domains are highly conserved

Retrieving all the predicted proteins annotated as containing the CAT-like domain (IPR023213) from the non-redundant database of predicted proteins from all seven tissues produced a list of 130 entries. Of these predicted proteins, many were not long enough to constitute functional BAHD acyltransferases, which are reported to have an average length of 445 amino acids and molecular masses ranging from 48 to 55 KDa (D'Auria, 2006). These short transcripts may be produced by truncated pseudogenes that have lost their function, or might be artefacts of the transcript assembly or the ORF prediction algorithms, such as spurious ORFs that are found nested within real ORFs. As these ORFs may have originated from incomplete transcripts, I would not have been able to design PCR primers to clone the complete coding regions of these genes. Yu et al. reported an artificially C-terminally

truncated transcript of a BAHD-AT that still encoded an enzyme (362 amino acids) with catalytic activity, although with a changed pH-optimum (Yu et al., 2008), but no naturally occurring truncated but functional BAHD-ATs have been described. I therefore decided to filter the predicted protein results by length, using a minimum cutoff of 400 amino acids, which reduced the list of entries to 77. This was done to reduce the complexity of the following analysis, but it is possible that true BAHD-ATs that may be relevant gene candidates were excluded in this process. The excluded predicted proteins may be revisited later to test additional candidate genes, if their complete coding sequences can be identified. I manually curated this list by removing entries that contained the CAT-like domain, but were not BAHD acyltransferases based on their homology to other enzymes, reducing the list down to 70 putative BAHD acyltransferases.

I constructed a phylogenetic tree of the putative grass pea BAHD acyltransferases using the EBI Multiple Sequence Comparison by Log-Expectation (MUSCLE) tool. This alignment is based on the amino acid sequences of the predicted proteins. This phylogenetic tree is displayed in Figure 64.
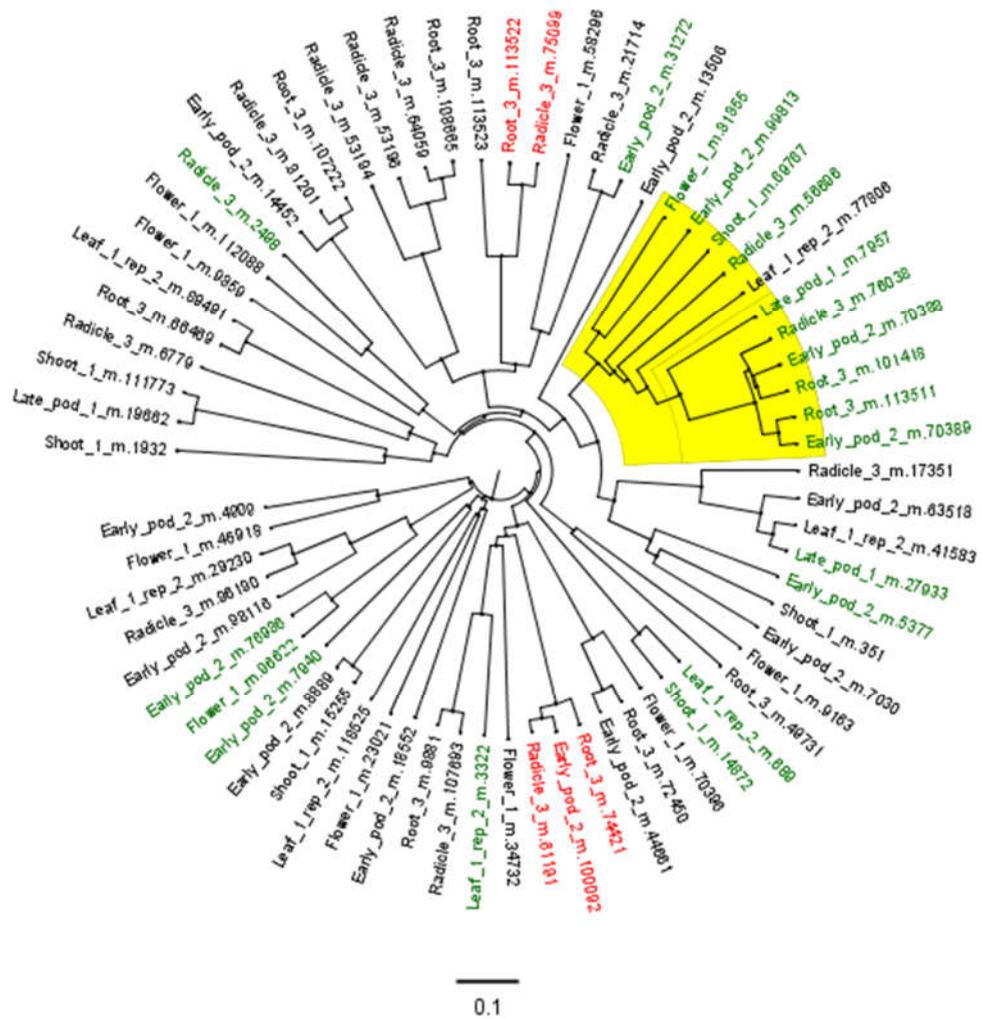
Figure 64. Phylogenetic relationship of putative BAHD-acyltransferases in grass pea based on alignment of amino acid sequences of predicted proteins. Gene models shown in green do not share HXXXDG domains with any close homolog in closely related species. Gene models shown in red contain disrupted HXXXDG domains and may be non-functional. The cluster highlighted in yellow was selected for biochemical testing for ODAP-synthase activity. Plotted using FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/)

4.3.6   Identification of ODAP-synthase candidates among the putative BAHD acyltransferases

Using the transcriptome data I applied the four characteristics described in section 4.1.2 to identify promising candidates for the ODAP synthase in grass pea:

(i)   The gene either does not have a close homologue in related species that do not produce ODAP, or differs in the amino acid sequence of the metabolically active HXXXD-domain

(ii) The gene is part of a cluster of closely related genes in grass pea

(iii) The gene has homology to known enzymes that catalyse N-acylation reactions

(iv) The gene's pattern of expression across tissues and developmental stages of the grass pea plant is correlated with the pattern of toxin distribution

To find the closest homologs of the putative grass pea BAHD-ATs, I searched predicted proteome databases of three related legume species, *Pisum sativum,* the closest relative for which an extensive sequence dataset exists and the two model legume species *Lotus japonicus* and *Medicago truncatula* using the blastp algorithm (Altschul et al., 1997)*.* The datasets used were the Pea RNA-Seq gene atlas hosted by the Institut National de la Recherche Agronomique (INRA) Dijon (Alves-Carvalho et al., 2015), the *L. japonicus* gene atlas hosted by the Noble Foundation (Verdier et al., 2013) and the *M. truncatula* transcriptome data hosted in the National Center for Biotechnology Information (NCBI) BLAST server, respectively. In addition, I performed searches including data from all species available on the NCBI server, excluding the three species already mentioned. In most cases, the closest homolog found in this search was from chickpea (*Cicer arietinum*).

Most of the putative grass pea BAHD-acyltransferases shared close homology with genes in these related species. Based on this homology, some of these putative BAHD-ATs could be ascribed likely functions. Two putative BAHD-ATs (the highly similar Root_3_m.113522 and Radicle_3_m.75099) did not contain a complete HXXXD-domain, but were missing one of the variable amino acids from the site. A mutation of this kind could be expected to disrupt the active site and lead to a non-functional enzyme. However, Sun et al. have recently described the *Anthocyanin Acyltransferase-Like* gene, which appears to have moved into the stem holoparasite *Cuscuta australis* via horizontal gene transfer from one of its legume hosts. Both in *C. australis* and in several legume species, the enzyme appears to have only a HXXD-type domain, while retaining enzymatic activity (Sun et al., 2016b).

Three highly similar putative BAHD-ATs (Root_3_m.74421, Radicle_3_m.61191 and Early_pod_2_m.100092) did not contain a histidine residue at the positions where the

HXXXD domain would be expected. Instead, these sequences contained an arginine residue at this position. While the histidine residue is generally considered necessary for functional BAHD-ATs, as it is catalytically active (Tuominen et al., 2011; Banks et al., 2011), there has been one case described of a BAHD-AT in which the histidine residue is replaced with a serine and is still catalytically active (Walker et al., 2002). Homologs of the three predicted proteins in *Pisum sativum*, *Medicago truncatula*, *Lotus japonicus* and *Cicer arietinum* all contained identical or similar RXXXDG sites. This degree of conservation raises the possibility that these genes still retain a function despite the lack of the histidine residue.

A total of 21 putative grass pea BAHD-ATs contained HXXXD-domains that were different from the HXXXD-domains in the majority of close homologs of these gene models in *Pisum sativum*, *Medicago truncatula*, *Lotus japonicus* and *Cicer arietinum* (shown in green in Figure 64). Most of these were spread across the radiation of BAHD-acyltransferases in grass pea and did not cluster. Due to the biochemical versatility of BAHD-acyltransferases it is unclear whether these genes represent novel biochemical functions or whether they share the same function of their homologs in other species. However, one group of these predicted proteins with HXXXD-domains unique to grass pea stood out (highlighted in yellow in Figure 64). In the following, these gene models are referred to using shorthand IDs, as shown in Table 21.

Table 21. Identity of gene models, associated transcripts and shorthand IDs for the members of the candidate clade of BAHD-acyltransferases highlighted in Figure 64.

Representative transcript	Gene model	ID
Shoot 1: c66479_g1_i2	Shoot_1_m.69767	BAHD 1
Flower 1: c76929_g1_i1	Flower_1_m.81855	BAHD 2
Early pod 2: c97996_g1_i1	Early_pod_2_m.99813	BAHD 3
Leaf 1 rep 2: c56878_g2_i1	Leaf_1_rep_2_m.77806	BAHD 4
Radicle 3: c58095_g1_i1	Radicle_3_m.76038	BAHD 5
Radicle 3: c54094_g1_i1	Radicle_3_56696	BAHD 6
Late pod 1: c15913_g1_i1	Late_pod_1_m.7957	BAHD 7
Root 3: c89105_g1_i1	Root_3_m.113511	BAHD 8
Early pod 2: c51463_g1_i2	Early_pod_2_m.70389	BAHD 9
Early pod 2: c51463_g1_i1	Early_pod_2_m.70388	BAHD 10
Root 3: c87597_g4_i3	Root_3_m.101418	BAHD 11

Of the 11 predicted proteins in this cluster, only one (BAHD 4) shares its HXXXD-domain with the predicted proteins of two of the other legume species (see Table 22). BAHD 1 and BAHD 3 each share HXXXD domains with their homologs in one related species (*P. sativum* and *L. japonicus*, respectively) but in both cases, the overall homology with is poor (<50% amino acid identity). BAHD 2 shares its HXXXD domain with its very similar (83% amino acid identity) *P. sativum* homolog. This radiation of genes with active sites different from their homologs in related species might suggest the evolution of a new biochemical function. The local alignment of the amino acid sequences of the eleven gene models surrounding the HXXXD-site is shown in Figure 65.

Figure 65. Local alignment of predicted peptide sequences of putative grass pea BAHD-acyltransferases from the ODAP synthase candidate clade. The environment of the catalytic HXXXD site is shown. The conserved domain is highlighted by a dashed red box.

Table 22. Summary of BLAST results comparing predicted grass pea proteins against homologs in other legume species. In each case, the HXXXD-domain of the grass pea predicted protein was compared to the HXXXD domains of the top ten BLAST hits.. BLAST scores, amino acid (AA) identities in %, and whether the HXXXD domains were identical or different are shown. * the closest pea homolog contained the same HXXXD domain as the predicted protein in grass pea, but the domain was not followed by a glycine.

	Pisum sativum			Medicago truncatula			Lotus japonicus			Cicer arietinum		
	BLAST score	AA identity	HXXXD domain	BLAST score	AA identity	HXXXD domain	BLAST score	AA identity	HXXXD domain	BLAST score	AA identity	HXXXD domain
BAHD 1	416	49	identical *	569	65	different	437	53	different	433	53	different
BAHD 2	783	83	identical *	608	65	different	465	53	different	569	65	different
BAHD 3	597	64	different	584	64	different	458	52	identical	578	64	different
BAHD 4	818	88	different	784	85	identical	555	61	different	692	76	identical
BAHD 5	787	87	different	674	73	different	569	64	different	657	72	different
BAHD 6	434	52	different	594	67	different	475	57	different	514	60	different
BAHD 7	818	87	different	580	64	different	499	58	different	619	69	different
BAHD 8	798	87	different	666	73	different	564	62	different	652	71	different
BAHD 9	783	87	different	668	72	different	552	62	different	645	71	different
BAHD 10	775	86	different	670	73	different	552	62	different	654	72	different
BAHD 11	808	89	different	677	73	different	556	63	different	660	73	different

As part of the transcriptome assembly pipeline, all predicted gene models in the transcriptome have been assigned automatic annotations based on sequence homology with proteins of known function, where possible. The entire candidate clade was annotated as "anthranilate N-hydroxycinnamoyl benzoyltransferase" (anthranilate N-HCBT), a gene first identified in *Dianthus caryophyllus* (Yang et al., 1997). The reaction catalysed by this enzyme bears noteworthy similarity to the proposed β-L-ODAP-forming reaction (Malathi et al., 1970), as both involve the acylation of a β-amino group of an amino acid as shown in Figure 66.
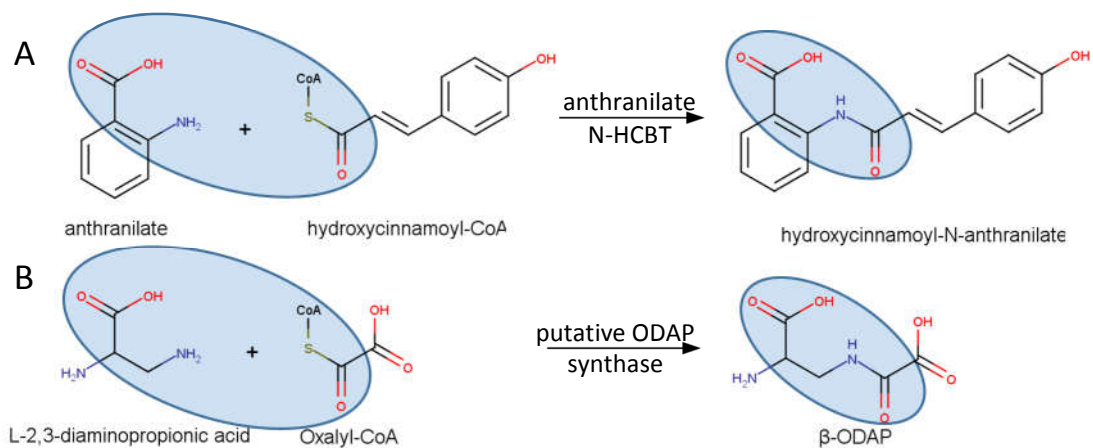
Figure 66. Simplified reaction schemes for the reactions catalysed by A) anthranilate N-hydroxycinnamoyl benzoyltransferase and B) the putative β-L-ODAP synthase. The similarity of the substrates and products surrounding the reaction centre is highlighted.

The similarity of these reactions suggests that only slight changes in the shape of the enzyme binding pocket might be necessary to change the substrate specificity to use the substrates of the β-L-ODAP-forming reaction. This circumstantial evidence strengthens the hypothesis that members of the candidate clade of BAHD-acyltransferases might be active as ODAP-synthases in grass pea. I therefore proceeded to test a subset of these genes biochemically by heterologous expression of their coding sequences in *Nicotiana benthamiana.*

4.3.7   Amplification of putative BAHD-acyltransferases from grass pea cDNA

The genes in the described clade of putative BAHD-acyltransferases represent the most promising candidates for enzymes with ODAP-synthase activity in grass pea. To optimise the annealing temperature for the amplification of candidate gene cDNA sequences from the reverse transcribed grass pea cDNA samples, I ran PCRs using two primer pairs (for BAHD3 and BAHD10/11, respectively, see section 4.2.8) using five different annealing temperatures ranging from 51.3°C to 59.3°C. I obtained amplified fragments of the

predicted size at all temperatures (data not shown). The melting temperatures of these regions are likely to be important only during the very first cycle of amplification, as in all following cycles the entire length of the oligo including the cloning sites are able to anneal to the amplified template, resulting in much higher melting temperatures. I therefore decided to use 52 °C as the annealing temperature for all amplifications.

I ran PCRs using the primer pairs described in section 4.2.8. The 5' and 3' ends of the BAHD10 and BAHD11 cDNA sequences and their 3'UTRs are identical, making it impossible to separate them at the amplification stage. As templates, I used cDNA derived from late pod tissues (BAHD3) or seedling shoot tips (BAHD8, BAHD9 and BAHD10/11) as these were the tissues showing the highest transcript abundances for these genes according to the read counting done using the RSEM algorithm (see section 4.2.9). The products of these PCRs were loaded on an electrophoresis gel (shown in Figure 67).
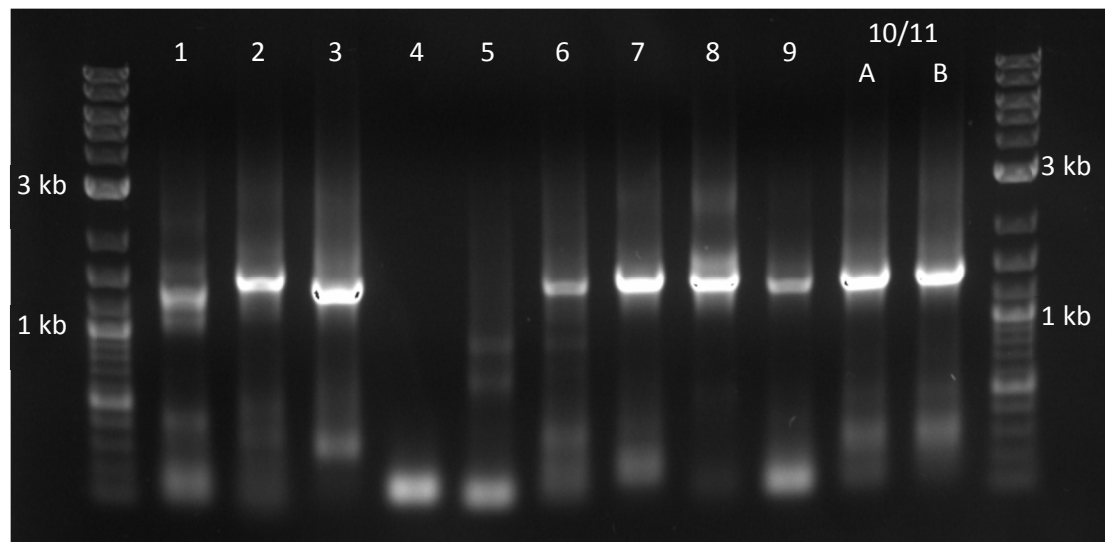
Figure 67. Electrophoresis gel showing PCR products of the members of the ODAP-synthase candidate clade. PCRs were run using gene-specific primers (see Table 10) and reverse transcribed cDNA from different grass pea tissues (see Table 11). Two reactions using different cDNA samples derived from shoot tip material were run for the pair of candidates 10 and 11 with identical primers. The gel is a 1 % w/v agarose TBE gel containing 0.01% v/v ethidium bromide. The leftmost and rightmost lanes contain DNA size markers, two of which are labelled.

Amplicons of the expected size (~1300 bp) were present for all candidate genes except BAHD4 and BAHD5. Time constraints forced me to prioritise and select the most likely candidates for the transient expression experiment. Out of the eleven predicted proteins in the clade, I thus selected five for testing experimentally for ODAP-synthase activity, based on their transcript abundance patterns shown in Table 23.

*Table 23. Transcript abundance of members of the BAHD-AT candidate clade across tissues in transcripts per million. Measured using the RSEM algorithm. * transcript counts for BAHD 9 and BAHD 10 were binned together as the two gene models are isoform transcripts of the same gene.*

	seedling shoot tip	seedling root tip	Leaf	root	flower	early pod	late pod
BAHD 1	3.13	5.02	0.40	0.12	0.18	0.57	1.98
BAHD 2	64.94	31.19	4.52	0.23	4.29	0.27	0.63
BAHD 3	647.07	551.13	527.10	681.35	609.75	696.87	857.53
BAHD 4	0.01	0.00	4.28	0.09	0.12	0.65	3.18
BAHD 5	2.44	2.81	1.28	1.83	1.67	2.25	1.55
BAHD 6	0.34	34.73	0.03	51.82	0.06	0.02	0.02
BAHD 7	0.00	0.00	1.21	0.14	4.82	9.40	60.10
BAHD 8	19.38	4.57	14.31	3.35	11.91	21.17	16.46
BAHD 9/ BAHD 10 *	76.07	10.73	46.08	3.02	29.18	60.27	39.23
BAHD 11	13.97	5.42	8.66	3.58	7.50	12.15	7.41

I chose BAHD 3 (Early_pod_2_m.99813) because of its high level of transcript abundance throughout all tissues of the plant, which far exceeded the abundance of any other transcript in the clade. I also chose BAHD 8, BAHD 9, BAHD 10 and BAHD 11 (Root_3_m.113511, Early_pod_2_m.70389, Early_pod_2_m.70388 and Root_3_m.101418 respectively) because the transcript abundance patterns of these genes most closely resembled the distribution of β-L-ODAP across the plant. It is possible that β-L-ODAP synthesis is catalysed by more than one enzyme and that these enzymes differ in their expression patterns across the plant. For example, one gene may be highly expressed in the developing pods of the plant and be responsible for the majority of ODAP produced there, while another might be highly expressed in young shoots and leaves. This would imply that a loss-of-function mutant in either of these genes would result in reduced toxin levels in only some tissues. However, among the three low-toxin mutants and the Indian low-toxin variety for which I measured ODAP levels using mass spectrometry (described in section 3.2.9), none showed differentially altered toxin levels compared to the high-ODAP parent. In all four low-toxin lines, the toxin levels were proportionately reduced in all tissues. All the low-toxin mutant lines that I selected from the mutant screen based on the ODAP levels in their seedling shoot tips and leaves, also produced low-ODAP seeds. This could imply that these mutations are in regulatory genes that affect β-L-ODAP synthesis across the plant, but some activity persists even if the regulatory gene is lost. Alternatively, they could represent partial loss-of-function mutations in genes encoding regulatory or metabolic enzymes. As has been observed in other mutant screens, most mutations caused

by EMS-mutagenesis only cause reduction in the activity of the gene product, rather than a complete loss-of-function (Takos et al., 2010). A third possibility is that the genes coding for metabolic enzymes involved in β-L-ODAP synthesis are redundant, and despite complete loss-of-function of the major gene, one or more other genes that code for enzymes that can catalyse the same reaction, albeit at a lower rate, persist.

The amplified cDNA of the selected candidate genes was purified by PEG precipitation. As shown by the electrophoresis gel depicted in Figure 68, PEG purification did not result in the complete removal of short (<500 bp) DNA fragments. For this reason, I further purified the PEG precipitated PCR products by running the remaining product on a separate 1% w/v agarose gel and performing gel extraction. No photograph was taken of this gel to avoid DNA degradation by UV-damage.
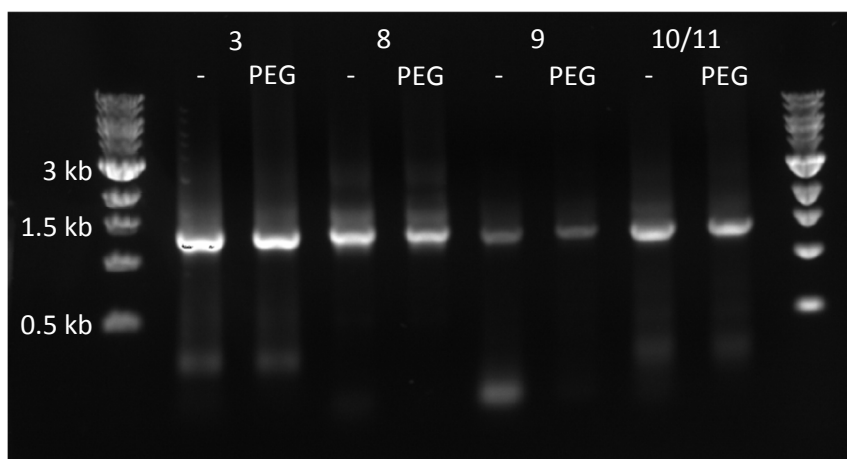
Figure 68. Electrophoresis gel showing the results of PEG precipitation of PCR products of ODAP synthase candidates. Even after PEG precipitation, short fragments persisted in the samples. The gel is a 1 % w/v agarose TBE gel containing 0.01% v/v ethidium bromide. The leftmost and rightmost lanes contain DNA size markers, three of which are labelled.

The purified candidate gene cDNA was cloned into donor vector pDONR207 using the Gateway® BP-cloning reaction (as described in section 4.2.11). I transformed the resulting entry clones into *E. coli* DH5α cells and grew up colonies overnight. I picked two colonies (A and B) of each BAHD3, BAHD8, BAHD9 and the BP reaction control and twelve colonies of BAHD10/11 (A-L) and performed colony PCRs using attL primers. I ran the products of these reactions on an electrophoresis gel (shown in Figure 69).
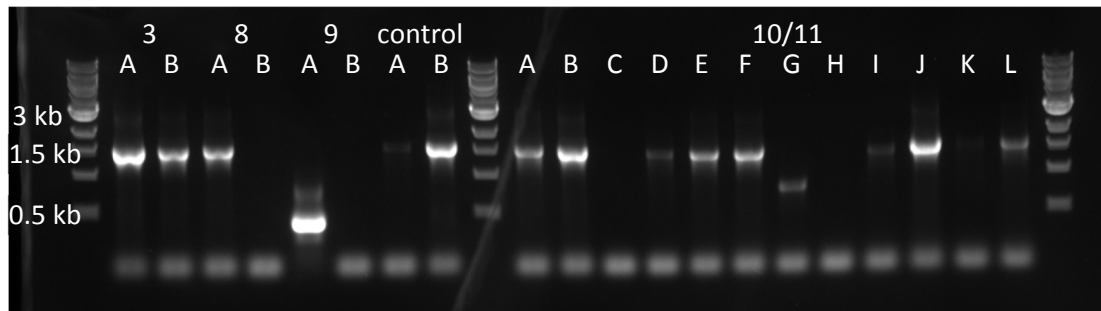
Figure 69. Electrophoresis gel showing products of colony PCRs of entry clones of the ODAP synthase candidates. PCRs were run using two E. coli colonies of transformants of BAHD3, BAHD8 and BAHD9, as well as the control for the BP reaction (pEXP7-tet, the tetracylce resistance gene and its promoter). PCRs were run using twelve colonies from BAHD10/11. The gel is a 1 % w/v agarose TBE gel containing 0.01 % v/v ethidium bromide. The leftmost, tenth and rightmost lanes contain DNA size markers, three of which are labelled.

At least one of the colony PCR for BAHD3, BAHD8, the BP reaction control and BAHD10/11 produced amplicons of the expected size, but the two colony PCRs for BAHD9 did not. I therefore picked four additional colonies of BAHD9 (C-F), as well as one BAHD8 (C) and four BAHD10/11 colonies (M-P) from the same transformation plates and performed colony PCRs (see Figure 70). Two of the BAHD9 colonies, the BAHD8 colony and the four BAHD10/11 colonies produced amplicons of the expected size (~1.3 kbp).
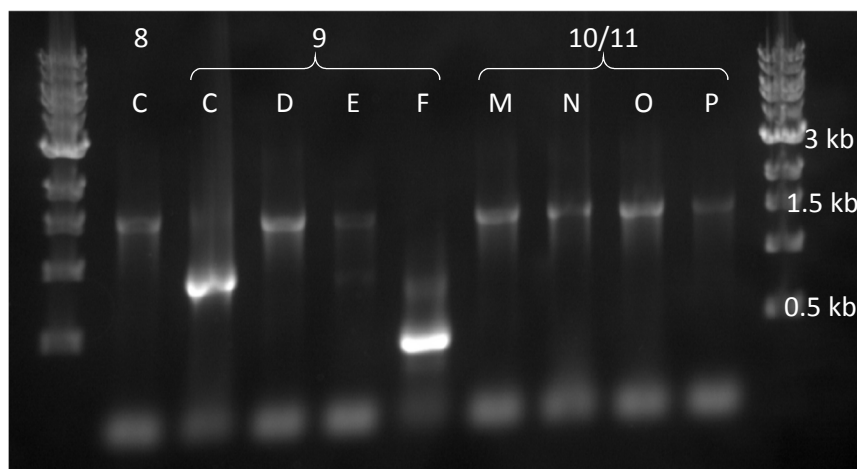
Figure 70. Electrophoresis gel showing products of colony PCRs of entry clones of the ODAP synthase candidates, run to replace failed colony PCRs shown in Figure 69. The gel is a 1 % w/v agarose TBE gel containing 0.01% v/v ethidium bromide. The leftmost and rightmost lanes contain DNA size markers, three of which are labelled.

To confirm the identity of the candidate coding sequences, I amplified and sequenced the inserts of the entry clones (discussed below). Based on the electrophoresis and sequencing results, I chose colonies 3A, 3B, 8A, 9D, 10/11B and 10/11J and inoculated liquid cultures to amplify these entry clones overnight. I performed plasmid minipreps to extract the amplified entry clones and used the extracted plasmids to perform the Gateway® LR-reaction as described in section 4.2.11. Transformants were spread on plates and incubated overnight. I picked colonies to perform colony PCRs, the products of which are shown in Figure 71. Colonies BAHD 3 AA, BAHD 3 BA, BAHD 8 AA, BAHD 9 DA and BAHD 10/11 JB produced amplicons of the expected size (~1.3 kb). I used these colonies to inoculate liquid cultures to amplify the expression clones.
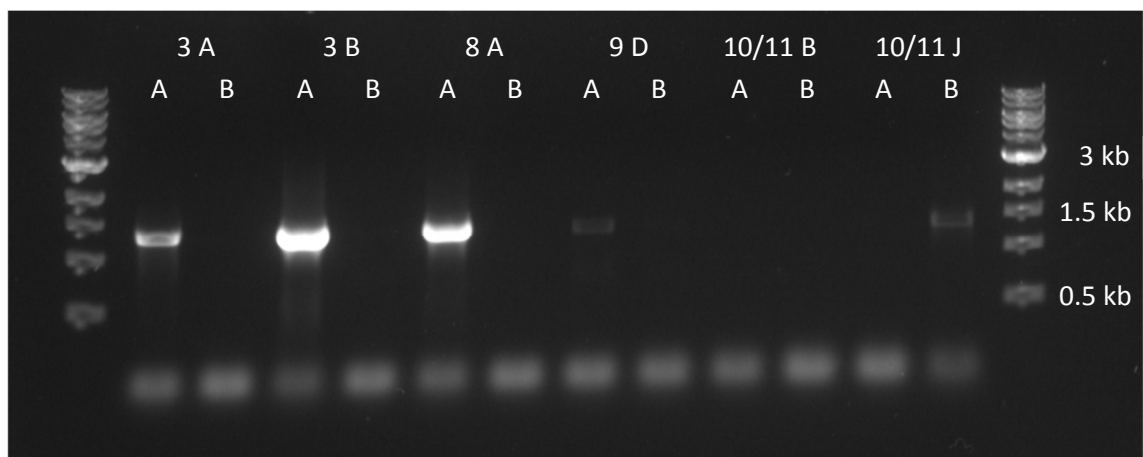
Figure 71. Electrophoresis gel showing the products of colony PCRs of E. coli colonies transformed with pEAQ-HT expression clones of the ODAP synthase candidate genes. PCRs were run using attB-primers. The leftmost and rightmost lanes contain DNA size markers, three of which are labelled.

To confirm the identity of the candidate coding sequences, I purified the products of the colony-PCRs and prepared them for sequencing. By aligning the sequences of entry clones and expression clones with the transcript sequences obtained from the transcriptome, I was able to check whether the correct coding sequence had been amplified and whether any errors had been introduced during the amplification. The sequence of the inserts in both entry and expression clones generated using the primer pair for BAHD 3 turned out to be identical to the expected transcript (Early_pod_2_c97996_g1_i1) in the transcriptome. The clones generated using both the primer pairs for BAHD 8 and BAHD 9 carried very similar inserts. Both of these matched most closely to the BAHD 9 transcript (Early_pod_2_c51463_g1_i2), but each carried several differences (BAHD 8: 40 single nucleotide polymorphisms, one deletion and one insertion each of six nucleotides, BAHD 9 clone 42 nucleotide polymorphisms, and one insertion of 6 nucleotides) compared to the

coding sequence of the BAHD 9 transcript. The sequences of the two clones also differed from each other in 13 nucleotides. No other sequence in the transcriptome aligned better to the coding sequences of the two clones than the BAHD 9 transcript. None of these differences introduced stop codons or shifts in the reading frame, but many resulted in amino acid changes (BAHD 8 clone: 23 amino acids substituted, 2 inserted, 2 deleted; BAHD 9 clone: 23 amino acids substituted, 2 inserted). In both clones, the coding sequences of the expression clones were identical to the coding sequences of the entry clones, but differed from the sequence found in the transcriptome. There are several possibilities for how these differences might have arisen.

(i)   Errors may have occurred during the reverse transcription of RNA and PCR amplification of these candidate genes from cDNA. However, 29 of the mismatches with the BAHD 9 sequence and one of the 6bp insertion/deletions are shared between the two clones, which were obtained using different primer pairs. It is highly unlikely that so many amplification errors would occur in parallel in two separate PCR reactions.

(ii)   The transcript sequence may contain errors due to low sequence coverage, meaning the cloned sequences might provide a more accurate reflection of the true sequence found in the plant.

(iii)   Highly similar transcripts produced by paralogs of the same gene may have given rise to assembly errors during the construction of the transcriptome, meaning sequences of non-identical transcripts have been combined. In this case, either or both of the cloned sequences may be true coding sequences.

(iv)   The same variety (LSWT11) was used for both transcriptome sequencing and cloning from cDNA, but genetic differences may have been present between the plants used for each application. The cDNA of BAHD3 was obtained from the same late pod RNA sample as used for the sequencing, but the cDNAs of BAHD8, BAHD9 and BAHD10/11 were obtained from a different young shoot tip RNA sample than the one used for transcriptome sequencing, because the original sample had become too degraded for successful cDNA amplification.

If at least some of the observed mismatches with the expected sequence were due to amplification errors, the structure and/or function of the encoded enzyme may be disrupted. However, none of these differences concerned the HXXXD domain or led to the truncation or frame-shift of the peptide sequence. This makes it impossible to tell from sequence data alone whether the function of the enzyme would be affected by these changes. I therefore decided to carry on with the experiment without excluding the BAHD 8 and BAHD 9 clones, but with the caveat that these may not represent the transcript found in grass pea. Hence a negative result for the catalytic activity of the enzymes encoded by these clones may not be conclusive.

The transcripts for BAHD 10 and BAHD 11 contained identical sequences at the beginning and end of their coding sequences and in the untranslated regions immediately upstream and downstream from it. This made it impossible to differentiate between the two coding sequences by the use of gene-specific primers. For this reason I used only one primer pair, aiming to amplify both coding sequences. After generating entry clones I then proceeded to sequence the inserts of plasmids from 15 individual *E.coli* colonies. As each bacterial colony was likely to derive from a single transformation event, I hoped that this way I would be able to capture at least one clone containing each of the two coding sequences. Sequences derived from 13 of these colonies aligned to transcripts in the transcriptome, while two produced poor quality sequences that did not align. All 13 of the alignable sequences aligned to the coding sequence of BAHD 10. All of these sequences contained the six codon exchanges that I intentionally introduced through my primer design to access a better gene specific primer region. All sequences also contained a 6 bp in-frame deletion compared to the BAHD 10 transcript, at the same sequence position (97-102) as the 6 bp insertion/deletion polymorphism between the cloned BAHD 8 and BAHD 9 sequences and the BAHD 9 transcript (see predicted protein structures in Appendix 3.7). I selected the colony that gave rise to the sample with the highest sequencing quality for further experiments. Apart from the intentional codon exchange and the 6bp deletion, the cloned BAHD 10 coding sequence was identical to the sequence obtained from the transcriptome.

In summary, the cloned BAHD 3 coding sequence proved to be identical to the transcriptome sequence. The cloned coding sequences for BAHD 8 and BAHD 9 were both most similar to the BAHD 9 transcript and contained many mismatches implying amplification errors and/or chimeras generated in the sequencing or assembly of the transcriptome. The cloned BAHD 10 coding sequence was identical to the expected sequence with the exception of a 6bp deletion near the 5'-end of the coding sequence. None of the clones contained the coding sequence of BAHD 11. Due to time constraints I decided to proceed with these four expression clones rather than designing alternative primer pairs to try to ascertain whether the source of the mismatches lies with the amplification and cloning or with the transcriptome sequencing.

I transformed *Agrobacterium tumefaciens* GV3101 cells with the expression clones or the pEAQ-HT-GFP plasmid, using electroporation as described in section 4.2.12. Colonies of transformants were used to inoculate liquid cultures to grow up bacteria for agroinfiltration. To confirm that the cultures contain the expression clones I ran colony-PCRs, the products of which are shown in Figure 72. All reactions using the candidate gene

expression clones produced amplicons of the expected size (~1.3 kb). The cultures carrying the pEAQ-HT-GFP did not produce any amplicons. This was expected, because the plasmid does not contain *attB* sites, which were used as primer regions. Based on these results, I selected the cultures BAHD 3 BAA, BAHD 8 AAA, BAHD 9 DAA and BAHD 10 JBA for agroinfiltration.
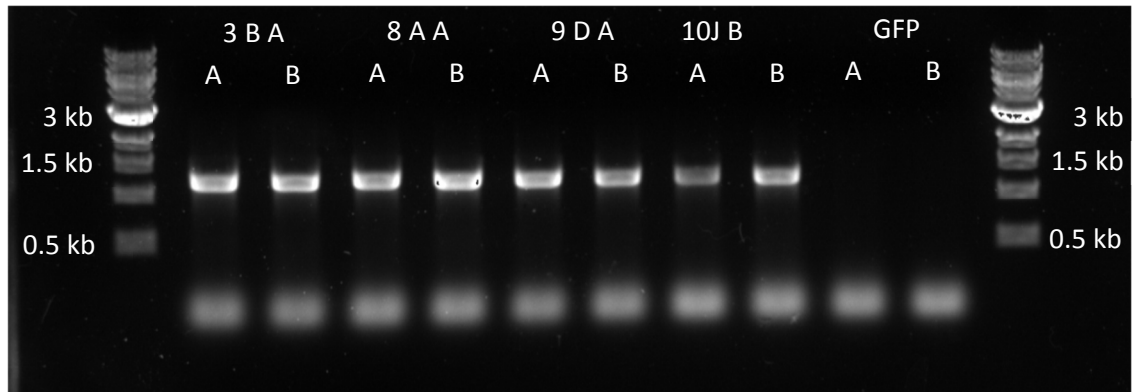
Figure 72. Electrophoresis gel showing the products of colony PCRs prepared from A. tumefaciens liquid cultures carrying the ODAP-synthase candidate expression clones. PCRs were run using attB-primers. The pEAQ-HT-GFP plasmid, which was used as a control, does not contain attB sites and was included here as a negative control. The leftmost and rightmost lanes contain DNA size markers, three of which are labelled.

4.3.8 Transient expression of ODAP-synthase candidates in *N. benthamiana*

I infiltrated *N. benthamiana* leaves with *Agrobacterium tumefaciens* suspensions carrying the expression clones of candidate BAHD-ATs, as described in section 4.2.13. Three days after the agroinfiltration, small lesions were visible in some of the infiltrated leaves (including the mock infiltrations), due to pressure damage at the injection site and 'ballooning' of tissue, as shown in Figure 73.

Pressure damage at infiltration site

'ballooning'
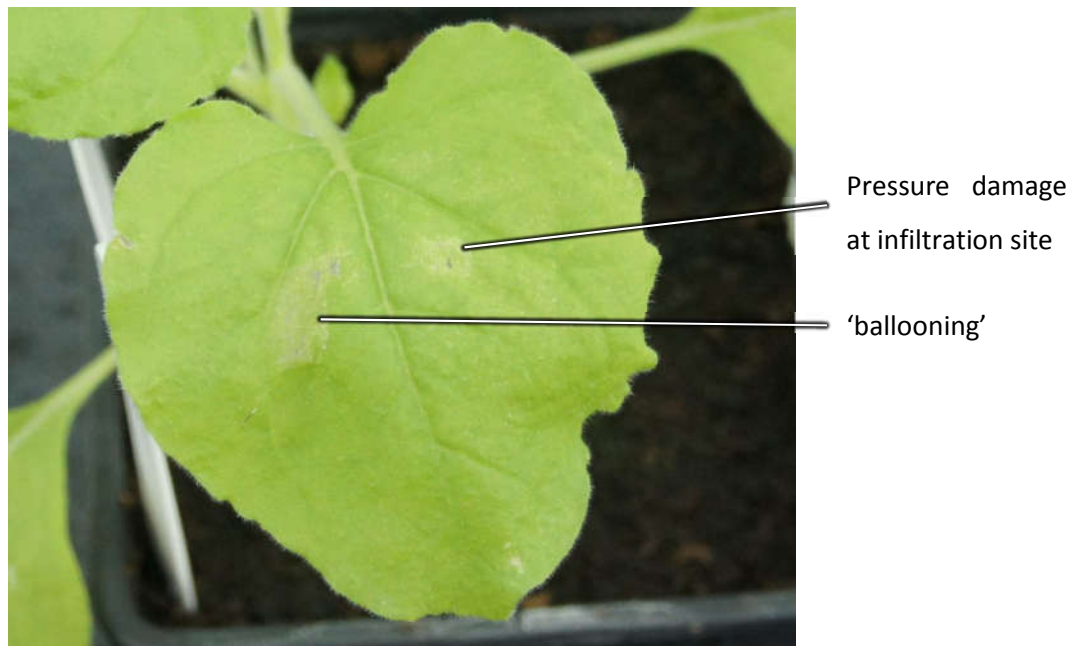
Figure 73. N. benthamiana leaf showing lesions three days after agroinfiltration. In some cases, the infiltration caused internal tissue rupture in mock infiltrations, agroinfiltrations and injections with metabolite solutions, resulting in 'ballooning'. No other tissue damage was observed.

Some additional damage was caused by the second round of infiltration, injecting either metabolite solution or water (negative control). Apart from these lesions, all infiltrated leaves remained viable up to the point of harvest. In the leaves infiltrated with the pEAQ-HT-GFP vector, green fluorescence was visible in some mesophyll cells five days after agroinfiltration, showing that the infiltration resulted in expression in the plant tissue.

The four putative grass pea BAHD-acyltransferases that I transiently expressed in *N. benthamiana* were subject to three separate metabolite treatments. Three days after the agroinfiltration, leaves were infiltrated with solutions of either L-DAP by itself or both L-DAP and oxalic acid. The β-L-ODAP-forming reaction has been reported to require L-DAP and oxalyl-CoA as substrates (Malathi et al., 1970). Since oxalyl-CoA was unavailable commercially, I decided to rely on the endogenous breakdown pathway for oxalic acid via oxalyl-CoA, which exists in other dicots (Foster et al., 2012; 2016). As a negative control, I infiltrated leaves with sterile water only.

I extracted free amino acids from these leaf samples, derivatised the extracts and analysed these samples using LCMS as described in section 4.2.14. Figure 74 A shows the retention time of a derivatised β-L-ODAP standard sample. A peak at the same position was apparent in the chromatograms of both samples from leaves infiltrated with the BAHD 3 expression

vector and L-DAP and both samples from leaves infiltrated with the same expression vector and L-DAP as well as oxalic acid (see Figure 74 C and E). This peak was absent in all other samples. In both standard and BAHD 3 samples, the peak was present for all four mass transitions that were measured, confirming that the peak does indeed represent β-L-ODAP. The transition 347.1 → 171.1 had the highest intensity, as I had observed before (see section 3.2.9).

No β-L-ODAP was present in leaves that had been mock infiltrated or agroinfiltrated with the GFP-vector (see Figure 74 B, D, F, H). No β-L-ODAP was present in leaves that had been mock-infiltrated or infiltrated with any of the candidate genes, but were not supplied with L-DAP (shown in Figure 74 G for BAHD3).

Substantial amounts of β-L-ODAP were produced in tissue expressing BAHD 3 in the presence of L-DAP as shown in Figure 74 C and E. This represented the first time that β-L-ODAP had been produced through heterologous gene expression and gave strong evidence that BAHD3 encodes an ODAP-synthase in grass pea. The high abundance of the transcript associated with BAHD3 in all the sequenced RNA samples indicates it is the most important ODAP-synthase in grass pea.

The other three expression clones did not result in the production of β-L-ODAP. As noted before, it is possible that amplification errors during the assembly of the BAHD 8 and BAHD 9 clones have rendered the encoded enzymes non-functional so further testing will be necessary to confirm that these genes do not code for ODAP-synthases in grass pea. The sequence of the BAHD 10 clone, however, corresponded well to the coding sequence in the transcriptome, except for one 6bp insertion/deletion (see Chapter 1App. 3.7). Despite its transcript abundance pattern that strongly resembled the distribution of β-L-ODAP across the grass pea plant, BAHD 10 does not appear to code for an ODAP-synthase, based on the transient expression experiment.

β-L-ODAP-production was also observed when both L-DAP and oxalic acid were added to the tissue. The additional presence of oxalic acid did not seem to increase β-L-ODAP production (compare Figure 74 E and C) and in fact I observed by far the highest β-L-ODAP content in a sample supplied with L-DAP alone (Figure 74 C), though a replicate of this treatment (BAHD3 x L-DAP) produced a peak of similar intensity to both replicates of the BAHD 3 expression with addition of L-DAP and oxalic acid treatment (not shown). This could indicate that sufficient oxalyl-CoA is present in *N. benthamiana* leaf tissue naturally or that exogenous oxalic acid is not effectively broken down via oxalyl-CoA.
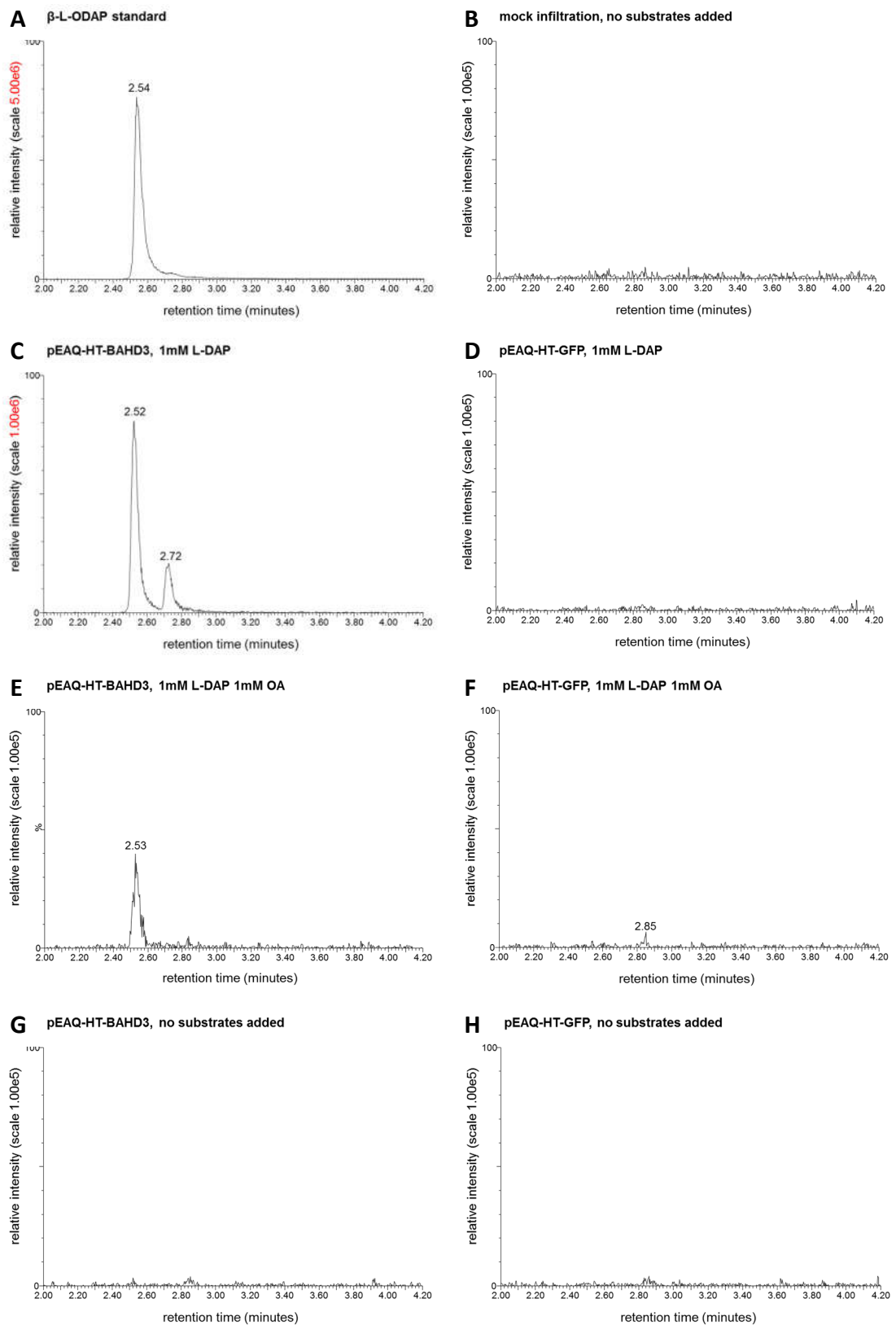
Figure 74. Chromatograms for the mass transition 347.1 → 171.1 showing the presence of β-ODAP in N.benthamiana samples transiently expressing the BAHD3 coding sequence, with addition of L-DAP (C) or L-DAP and oxalic acid (E). No β-ODAP was observed in negative controls (B,D,F,G,H) or in samples expressing other ODAP synthase candidates (not shown). In each case, 5 mg of freeze-dried leaf tissue were used (B-H). Intensities in A and C are shown at different scales than the other panels due to the high intensity of the peaks in A and C.

In the chromatogram of the *N. benthamiana* sample with the highest concentration of β-L-ODAP (Figure 74C) a second peak with a retention time of 2.72 minutes was observed. This peak was not present in the derivatised β-L-ODAP standard or in any of the derivatised *N. benthamiana* extracts. This second peak exhibited the same four mass transitions as the β-L-ODAP standard peak. This peak may represent an isomer of the derivatised β-L-ODAP, most likely derivatised α-ODAP. α-ODAP is reported to be present in grass pea, where it represents in the order of 5 % w/w of total ODAP (Roy and Rao, 1968). The two isomers are able to interconvert non-enzymatically (Abegaz et al., 1993; De Bruyn et al., 1994) and this process is accelerated in high or low pH environments. It is possible that the enzyme, itself, produces a small amount of α-ODAP due to imperfect substrate specificity or that the α-ODAP is subsequently formed through non-enzymatic conversion from β-L-ODAP.

4.4   Summary

The multi-tissue transcriptome of grass pea that was developed as part of this project represents a crucial step forward in researching this orphan crop. This new genetic resource will support the development of genetic markers to allow marker-assisted breeding of new grass pea varieties and comparative genomics to investigate the crop's remarkable tolerance to drought, flooding and some biotic stress factors. In addition it can be used to identify candidate genes based on biochemical knowledge and sequence homology. The power of this approach is underlined by the identification of a transcript coding for an enzyme with β-L-ODAP-forming activity from grass pea. By screening the transcriptome for genes encoding enzymes involved in related biochemical reactions, analysing their expression based on transcript abundance across the plant and comparing the sequences of grass pea transcripts to related species, I was able to identify transcript candidates for this enzyme, one of which I was able to confirm as a β-L-ODAP synthase through heterologous expression. This is the first enzyme in the β-L-ODAP biosynthetic pathway to be identified and sequenced. This discovery will allow us to greatly expand our knowledge of β-L-ODAP synthesis in grass pea and opens up a direct route to developing toxin-free genotypes of grass pea by screening for mutants in this gene and any homologs with β-L-ODAP-synthase activity or disrupting them through the use of genome editing.

Chapter 5 – General discussion

The objective of this study was to identify genotypes of grass pea with reduced levels of β-L-ODAP in their tissues and to understand better the biosynthesis of this neurotoxin. For this purpose I measured ODAP concentrations in seeds of collections of grass pea accessions, screened an EMS-mutagenised population to identify low-ODAP mutants and used transcriptome data to select candidates for the gene encoding ODAP-synthase, of which one was confirmed through heterologous expression.

5.1   Variation in β-L-ODAP levels among grass pea germplasm

For the analysis of germplasm presented in this thesis, I selected grass pea accessions from countries cultivating the crop around the world and screened a population of grass pea landraces from Ethiopia. The range of ODAP concentrations I observed was similar to previous germplasm studies that found seed β-L-ODAP concentrations to vary between 0.15 % and 0.75 % w/w. Low β-L-ODAP is commonly defined as being below 0.1 % w/w (Tay et al., 2000; Sharma et al., 2000; Campbell, 1997; Sarwar et al., 1995; Deshpande and Campbell, 1992). Most of these studies only investigated the variation in β-L-ODAP levels in grass pea accessions from individual countries. To allow better screening of the available genetic diversity of grass pea for β-L-ODAP levels as well as other traits, it is necessary to establish a core collection that includes a limited number (in the order of hundreds) of accessions representing the widest possible sample of the genetic diversity of *Lathyrus sativus* (Vaz Patto and Rubiales, 2014). I attempted to achieve a more widely spread genetic diversity by selecting accessions based on diverse geographic origins, but this approach could be improved by selecting lines based on genetic markers. Such a genetic screening could build on recent work by Wang et al. (2015), but would require a larger set of accessions to be screened to establish a core collection of global grass pea diversity, since theirs only included 266 *Lathyrus sativus* accessions, with few of them sampled from South Asia.

The plate-based assay I developed represents a significant improvement on the previously described spectrophotometric method for assaying large numbers of samples. My screen measuring ODAP in seeds of international grass pea accessions showed wide variation in

seed ODAP concentrations, ranging from 0.14 % to 0.76 % w/w of dry weight (not including the low-ODAP variety LS8246 with an ODAP concentration of 0.04 % w/w), but did not reveal any geographical patterns. The strength of this analysis is limited by the relatively low number of accessions screened (140 accessions, of which 52 were split into sub-accessions), possible issues of contamination and the fact that both landraces and cultivars were included. The analysis of Ethiopian landraces similarly did not reveal any correlation between ODAP concentration and geographical or altitudinal origin of the collection site.

All accessions that showed low (< 0.1 % w/w) seed ODAP concentrations were known low-ODAP accessions from Australian, Canadian and Indian breeding programmes. One accession among the population of Ethiopian landraces that had not been screened previously showed intermediate seed ODAP concentration (~0.14 % w/w), but no new low-ODAP accessions were discovered in any of the three germplasm collections. Several low-ODAP grass pea varieties have been developed by selection from landraces in the past (Campbell and Briggs, 1987; Dixit et al., 2016), but many of these low-ODAP varieties do not show genetic complementation when crossed, indicating that their low-ODAP phenotypes are due to alleles in the same gene or genes (Campbell, 1997). Attempts have been made to widen the gene-pool by crossing *Lathyrus sativus* with other *Lathyrus* species, some of which do not produce β-L-ODAP. *Lathyrus sativus* does not readily hybridise with other species in the genus (Kumar et al., 2011b), but crosses with *Lathyrus pseudocicera* have given rise to viable plants. Crosses with *Lathyrus cicera* could be recovered through the use of embryo-rescue. However, in neither case did the offspring of these crosses segregate for β-L-ODAP content (Addis and Narayan, 2000). The rarity of low-ODAP accessions and the difficulty of doing inter-species crosses imply that it may not be possible to decrease β-L-ODAP levels in grass pea further using existing germplasm alone.

5.2   Comparison of low-ODAP induced mutant lines and released low-ODAP varieties

By further adapting the plate-based spectrophotometric assay for the screening of large numbers of samples, I developed a high-throughput assay that I used to screen an EMS-mutagenised population of grass pea for low-ODAP mutants, as described in Chapter 3. Three of the low-ODAP mutant lines I identified in the EMS-mutant screen showed lower

seed β-L-ODAP concentrations than the popular Indian low-ODAP variety Mahateora as measured using the spectrophotometric assay and/or the LCMS-method using the heavy-isotope labelled internal standard. The levels of β-L-ODAP in these mutant lines compare favourably with low-ODAP grass pea varieties that have been released to date, such as LS8246, Prateek, Mahateora, Ratan and Wasie (Siddique et al., 2006; Kumar et al., 2011a; Campbell and Briggs, 1987), but comparison across multiple environments, specifically drought stress and zinc deficient soils, will be necessary to ascertain whether the low ODAP levels in these mutants are stable (Polignano et al., 2009). The mutant alleles contained in these lines will make a useful contribution to breeding programmes for new low-ODAP varieties, especially if low-ODAP traits caused by separate genes can be combined. This approach of accumulation of low-ODAP alleles may be helpful in reducing the risk of β-L-ODAP-production increasing to dangerous levels under intense environmental stress (Fikre et al., 2008).

During the second and third passes of the mutant screen and during the subsequent confirmation stages, I excluded several M2 families that showed intermediate ODAP concentrations. Some of these may have been false positives in the screen that do not contain lower ODAP concentrations than the wild type, but others may contain mutants with intermediate ODAP concentrations. These could represent mutations in redundant genes that contribute a minor part of the enzymatic activity at one step of ODAP biosynthesis, regulatory genes that are not essential for ODAP synthesis or reduced-activity mutations in major genes. Any of these would be useful additions to the set of low-ODAP mutants I included in my complementation analysis; the former set of mutations may represent additional complementation groups, while the latter may add to the allelic series in the complementation groups I observed. To re-test these M2 families, new seeds from the original M2 seed packets would have to be sown and assayed to identify new mutant plants, and confirmed by assaying the ODAP concentration in seeds they produce. In this way, it may be possible to identify several more low-ODAP mutant lines without having to conduct another mutant screen.

5.3   Crossing low-ODAP genotypes

To analyse gene complementation among the low-ODAP mutants, I performed a total of 671 crosses between different low-ODAP mutants, of which 250 produced seeds. The

preliminary results of these crosses revealed between two and five complementation groups among the mutant lines indicating that mutations in at least this many genes may cause the low-ODAP phenotype. Some or all of these may encode enzymes that catalyse steps in the biosynthesis of β-L-ODAP or regulatory gene products. These mutants will provide a useful resource for investigating the production of β-L-ODAP as well as the steps of its biosynthesis and could be used to further reduce β-L-ODAP contents by developing double mutants. Among the crosses between low-ODAP genotypes (i.e. low-ODAP mutants and Mahateora) and comparatively high-ODAP genotypes (i.e. LSWT11 and Nirmal), I observed a maternal effect concerning the β-L-ODAP levels in seeds and young seedlings. Maternal effects on the β-L-ODAP concentrations in seeds (such as depositional phenomena or cytoplasmic inheritance) had been suggested as possibilities in previous reports (Campbell, 1997; Quader et al., 1987; Tiwari and Campbell, 1996), but have not been further elucidated. When re-testing plants at a later developmental stage, the maternal effect disappeared, suggesting that it was due to accumulation of β-L-ODAP in the seed by the maternal parent, rather than cytoplasmic inheritance. Because of the maternal effect affecting β-L-ODAP levels in seeds and seedlings and the problem of high background readings when testing leaf tissues as described in section 3.3.11, I have not yet been able to assign all complementation groups with certainty. To confirm these data the ODAP in F2 seeds produced by these F1 plants must be measured. These results will confirm or refine the complementation analysis I made based on leaf samples. Depending on these results further crosses between the low-ODAP mutant lines may need to be performed to assign all mutant lines and the low-ODAP variety Mahateora to complementation groups with greater confidence.

The F2 seeds collected from the F1 plants have resulted from self-fertilisation. The seedlings that germinate from these seeds can be expected to segregate for the mutant alleles inherited from the homozygous mutant lines. F2 plants that have originated from crosses of mutants in different complementation groups will be assayed for β-L-ODAP to test if the levels of the toxin can be further reduced in homozygous double mutants.

Once the complementation groups of the 14 mutant lines that I have crossed are confirmed, representatives from each complementation group can be crossed with newly identified low-ODAP mutants from previously excluded M2 families to expand the allelic series of the complementation groups and potentially identify mutants that fall into additional complementation groups. The same could be done for existing medium- and low-ODAP varieties. Several of these (LS8246, Ratan and Prateek) derive from the same

low-ODAP germplasm as Mahateora, i.e. the medium-ODAP variety Pusa-24. These are likely to fall into the same complementation group as Mahateora. Therefore crosses with low-ODAP varieties derived from different germplasm, such as Ceora from Australia (Siddique et al., 2006) or Wasie from Ethiopia (Kumar et al., 2011a), would be more informative.

5.4   A potential oxalyl-CoA synthetase revealed by the grass pea transcriptomes

I extracted RNA from seven samples of LSWT11 tissue covering various developmental stages and organs (as described in Chapter 4). These RNA samples were sequenced to generate transcriptomes by de novo assembly. One of the transcripts identified from these transcriptomes shared close sequence homology with genes encoding oxalyl-CoA synthetases in *Arabidopsis thaliana,* and *Medicago truncatula* (Foster et al., 2012; 2016). The enzyme encoded by this transcript is currently being investigated through heterologous expression in *E. coli* (personal communication, Anne Edwards, JIC). The formation of oxalyl-CoA represents the penultimate reaction step in the biosynthesis of β-L-ODAP. If this enzyme can be proven to have oxalyl-CoA synthetase activity, it would be very likely that it is involved in the β-L-ODAP-biosynthesis pathway in grass pea and may be the same enzyme as the oxalyl-CoA synthetase that was partially purified by Malathi et al. (Malathi et al., 1968; 1970). Experiments to test the biochemical function of this enzyme in vitro and by heterologous expression are currently being conducted. Once confirmed, it will be necessary to knock down its activity in grass pea (using RNAi or Virus Induced Gene Silencing) to see if this affects oxalate breakdown and β-L-ODAP content in seeds and other tissues.

If there is no parallel pathway for the synthesis of oxalyl-CoA in grass pea, knocking out the gene encoding this enzyme and any of its isoforms may even yield plants unable to synthesise β-L-ODAP. However, this might prove to be only of limited usefulness for the development of new, safe varieties, because the plant's ability to break down oxalic acid effectively would also be impaired. Both increased susceptibility to *Sclerotinia sclerotiorum* and greatly reduced germination rates have been observed in Arabidopsis a*cyl-activating enzyme 3* (At*aee3*) T-DNA insertion lines (Foster et al., 2012). Equally, *Medicago truncatula Mtaae3* RNAi knock-down lines showed increased susceptibility to *S. sclerotiorum* and increased calcium oxalate crystal formation. Seed germination was not compromised in

these lines, which Foster et al. argued may be due to the knock-down of *Mtaae3* being incomplete (Foster et al., 2016). Grass pea plants that were entirely unable to break down oxalic acid in this way might suffer from greatly increased susceptibility to necrotrophic pathogens that secrete oxalic acid and potentially exhibit defects in seed development as seen in *Arabidopsis aee3*-mutants (Foster et al., 2012). In addition, increased levels of oxalate would be highly unwelcome in a crop variety, as oxalate is liable to lock up calcium and other metal ions in nutritionally inaccessible crystals. Oxalate may also lead to the development of stones in the urinary tract of mammals (Noonan and Savage, 1999), making high-oxalate grass pea varieties potentially unsuitable as food or feed crops. For these reasons, grass pea genotypes lacking the oxalyl-CoA synthetase may not be a useful breeding goal in themselves.

Nevertheless, these disadvantages might be mitigated by introducing a novel catabolic pathway for oxalic acid into a grass pea genotype in which oxalyl-CoA synthetase activity has been knocked out. Recently, Kumar et al. succeeded in expressing the oxalate decarboxylase enzyme isolated from the Enoki mushroom (*Flammulina velutipes*) in a high toxin (~0.7 % w/w) grass pea genotype, through Agrobacterium-mediated stable transformation (Kumar et al., 2016). This resulted in 75 % reduction in seed oxalic acid levels and 73 % reduction in seed β-L-ODAP levels, improved tolerance to *S. sclerotiorum* and increased bioavailability of calcium, magnesium, iron, zinc and manganese (Kumar et al., 2016). The transgenic lines the authors produced would not be classed as 'low-toxin' varieties, as they still contained between 0.2 % and 0.35 % w/w β-L-ODAP in their seeds, but this remains an important proof of principle, as it shows that oxalate can be diverted into a different breakdown pathway, reducing β-L-ODAP production.

Another pathway for oxalate breakdown via oxidation has been described in monocots. Oxalate oxidase activity appears to be rare in dicots (Membre et al., 1997; Rietz et al., 2012), but has been recently observed in azalea (*Rhododendron mucronatum* G. Don) (Sakamoto et al., 2015). Transgenic expression of the oxalate oxidase enzyme from barley (*Hordeum vulgare*) in peanut (*Arachis hypogea*) resulted in increased tolerance to externally applied oxalic acid and reduced susceptibility to the pathogen *Sclerotinia minor* (Livingstone et al., 2005). A transgenic grass pea line expressing either the fungal oxalate decarboxylase or the monocot oxalate oxidase in the background of an oxalyl-CoA-synthetase knockout might be unable to produce β-L-ODAP, while still managing to maintain appropriate levels of oxalate. With either transgene, some fine-tuning may be needed to achieve levels of oxalate in the plant that retain its useful functions in terms of

defence against herbivory and ion tolerance, without compromising the nutritional quality of the crop.

A drawback of this strategy, however, is that any variety produced this way would be classed as a genetically modified organism (GMO). Despite the potential of GM-technologies for the improvement of crops for low-income countries, and generally positive effects for economic growth benefitting the poor where they have been introduced (Thirtle et al., 2003; Krishna and Qaim, 2008; Carpenter, 2010), serious concerns exist among some consumers, farmers and non-governmental organisations that have resulted in legal restrictions on the cultivation of such crops (Singh et al., 2006; Purchase, 2005). In the current regulatory climate regarding transgenic crops, this would cause delays and increase the costs associated with the introduction of new low/zero-ODAP varieties in target countries (Bett et al., 2010), and may have unforeseen negative effects on perception by the public (Purchase, 2005). An example of this issue is the slow introduction of 'Golden Rice'. This publicly-funded research project aimed to develop rice varieties that accumulate provitamin A in the endosperm in order to counter vitamin A deficiency in low- and middle-income countries. Because the desired trait was not present within the rice genepool, the researchers used a transgenic approach and succeeded in developing genotypes with greatly increased provitamin A content. However, the release of these varieties has been delayed by at least 16 years due to complex regulatory hurdles facing the introduction of GMOs (Potrykus, 2010; Wesseler and Zilberman, 2016). While transgenic approaches may be an interesting route of investigation to enhance our understanding of oxalate metabolism and β-L-ODAP synthesis, bringing GMO grass pea varieties to the field would likely be far more expensive and time-consuming than non-GM varieties due to these legal regulations.

5.5   Investigating the grass pea ODAP-synthase

To identify candidate genes for the final step of β-L-ODAP-synthesis I extracted RNA and sequenced transcriptomes from several grass pea tissues. The biochemistry of the synthesis reaction, in which a CoA-derivative acylates an amino group (Malathi et al., 1970; Ghosh et al., 2015), pointed to a BAHD-acyltransferase as the catalyst (Bontpart et al., 2015; D'Auria, 2006). By comparing grass pea BAHD-ATs to sequences of related legume species, I was able to identify predicted proteins with potentially novel functions (see section 4.3.6). Using homology-based annotation and transcript abundance data, I decided on a subset of

grass pea BAHD-ATs which I investigated biochemically. I cloned these genes and expressed them heterologously in *Nicotiana benthamiana*. I decided to supply L-DAP as an intermediate, because it had previously been shown to act as a substrate to a β-L-ODAP forming reaction in vitro (Malathi et al., 1970; Ghosh et al., 2015). This experiment confirmed one grass pea transcript as encoding an enzyme that is capable of synthesising the formation of β-L-ODAP in *Nicotiana benthamiana* leaves. Because of the high abundance of this transcript across grass pea tissues it is likely that this transcript encodes the main β-L-ODAP synthase in grass pea and may be the same enzyme as the one partially purified previously (Malathi et al., 1970; Ghosh et al., 2015). Small amounts of α-ODAP were produced as by-products of this reaction, although it cannot be ruled out that it may have been generated through later isomerisation in vivo or during the extraction procedure.

The activity of this enzyme needs to be characterised in in vitro experiments. This could be achieved by extraction of the enzyme from grass pea tissue (Malathi et al., 1970; Ghosh et al., 2015), or by extraction from *N. benthamiana* or a bacterial culture expressing the cloned gene. To produce purified enzyme, it may be helpful to add an N-terminal or C-terminal amino acid tag to the enzyme sequence to facilitate selection. N-terminal His-tags (Berger et al., 2006; D'Auria et al., 2007b; Luo et al., 2009; Kosma et al., 2012) have been successfully used to isolate functional BAHD-ATs, but it should be noted that some enzymes of this family appear to have greatly reduced activity when tagged with N-terminal or C-terminal epitopes (D'Auria et al., 2002).

To further characterise the grass pea ODAP-synthase it will be necessary to identify the genomic sequence, which encodes the enzyme. This will hopefully be achieved by interrogating the grass pea draft genome that is currently being assembled by JIC in collaboration with the Earlham Institute using Illumina whole genome shotgun sequencing (WGS), following the Discovar pipeline (Weisenfeld et al., 2014). The scaffold assembly will be refined by optical mapping using the Irys system (BioNano Genomics, San Diego, California, USA) (Shelton et al., 2015). This genome is based on the variety LS007. The genome sequences of the model legumes *Medicago truncatula* and *Lotus japonicus* are being used to assemble the genome scaffolds. Once it becomes available, the *Pisum sativum* genome, which is currently being prepared for publication by the international consortium for Pea Genome Sequencing (personal communication from Judith Burstin, INRA, Dijon, France) will be used to improve the grass pea assembly. *Pisum sativum* is

closely related to grass pea and has a genome of comparable size (Greilhuber and Ebert, 1994), which may be particularly useful in the assembly of highly repetitive regions.

Alternatively, the gene sequence could be identified by using inverse PCR to amplify the genomic region surrounding the cDNA and sequence it (Ochman et al., 1988). Knowledge of the gene sequence surrounding the β-L-ODAP synthase will allow genetic markers to be identified that can be used to enable marker-assisted breeding of zero-/low-ODAP genotypes. This could help to accelerate future breeding efforts, but still relies on zero-/low-ODAP genotypes being identified or generated through other means, such as TILLING or genome editing.

TILLING (Targeting Induced Local Lesions IN Genomes) is a reverse genetics method to find mutations in known genes of interest. The technique relies on a repository of genomic DNA extracted from a mutant population, which can be interrogated for mutations. When the technique was first developed (McCallum et al., 2000; Colbert et al., 2001; Perry et al., 2003), this was achieved by allowing single strands of the mutant DNA, amplified from the region of interest, to anneal to single strands of DNA amplified from the reference genotype that was used for mutagenesis. This results in mismatches, typically involving a single base pair, at the loci of mutations. The heteroduplex DNA is then digested by an endonuclease that only targets single stranded DNA. This enzyme causes DNA to be fragmented where a mismatch is present, while duplex DNA without any mismatches is not digested. By separating the fragments resulting from each mutant according to their molecular weight, individuals in which this digestion has taken place, i.e. where a mutation in the target region exists, can be identified. More recently, advances in DNA-sequencing technology have enabled TILLING-by-sequencing (Tsai et al., 2011), which relies on sequencing of the amplicons of target genes from a population of mutants and comparison of the obtained sequencing with a reference sequence. TILLING-by-sequencing has advantages compared to TILLING by physical or enzymatic means, because in addition to identifying plants with mutations in the target gene, it also immediately reveals the exact sequence change caused by the mutation event and its effect on the amino acid sequence of the relevant gene product. TILLING could be used to rapidly screen a mutant population for mutations in the β-L-ODAP synthase gene, which I have identified, potentially revealing additional low-/zero-ODAP mutants. To this end, a TILLING population of grass pea based on the variety LS007 is currently being generated by the company BenchBio, Valvada, Gujarat, India, because the original mutant population used for the forward mutant screen (Chapter 3) did not have a sufficient mutation density to be suitable for TILLING.

An alternative approach to generating mutants in a target gene would be genome editing. This term refers to a number of techniques that change specific regions in a target DNA sequence, typically by targeting endonuclease activity to a pre-determined DNA-sequence. This has been achieved using meganucleases (Epinat et al., 2003), zinc-finger-nucleases (Bibikova et al., 2003; Miller et al., 2007) and transcription-activator-like-effector nucleases (TALENs) (Christian et al., 2010). The most recent technique uses sequence targeting by engineered bacterial clustered, regularly interspersed, short palindromic sequences (CRISPR) coupled with DNA cleavage by the Cas9 endonuclease (Hwang et al., 2013; Belhaj et al., 2013). The primary advantage of the CRISPR/Cas system compared to previous methods of genome editing is that the sequence targeting can be achieved by a synthesised guide RNA, rather than an engineered protein. This allows accurate targeting of any DNA sequence in a genome. The blunt-ended double-stranded break induced by the endonuclease can be used as the insertion point for a double-stranded DNA sequence (such as a transgene) with 5' and 3' ends homologous to the sequences adjacent to the double-stranded break. If no insertion sequence is supplied, the error-prone nature of DNA repair by non-homologous end joining may result in insertion/deletion mutations. These could cause disruption of genes, especially if they result in a reading frame shift. By targeting a conserved region in the genomic sequence of the ODAP-synthase, the CRISPR/Cas system could be used to create an allelic series of different levels of activity, including complete knockouts, of the ODAP-synthase gene. This would allow testing whether a complete knockout of this gene reduces β-L-ODAP production to zero and at the same time create alleles that enable the fine-tuning of β-L-ODAP production to study the physiological role of this compound in grass pea.

To enable genome editing in grass pea, both as a research tool and as a means to accelerate crop improvement, a CRISPR/Cas platform for grass pea will be developed (personal communication, Abhimanyu Sarkar, JIC). This will require the establishment of a robust transformation method for grass pea to introduce the CRISPR/Cas cassette consisting of the *Cas9* gene, a selectable marker and one or more guide RNAs. Methods for the transformation and in vitro regeneration of grass pea have been described, but show poor efficiency (Barik et al., 2005; Santha and Mehta, 2001). The recent publication that claimed successful transformation of grass pea with an oxalate decarboxylase gene (Kumar et al., 2016) does not state the genotype used and the described methodology was not successfully reproducible using LS007 (personal communication, Julia Russell, JIC). More efficient and reproducible methods for the transformation of grass pea and the

regeneration of plants will be necessary to allow the application of the CRISPR/Cas system in this species. This will allow targeted changes to be made in grass pea genes.

Genome editing using the CRISPR/Cas system requires transformation of the host organism with the genes encoding the guide RNA and the Cas9 endonuclease, but these genes can be removed through crossing, resulting in an organism containing the targeted mutation or the inserted sequence, but no other foreign genetic material (Xie and Yang, 2013). This raises the question of how crops that have been edited using the CRISPR/Cas system or other genome editing techniques but do not carry transgenic DNA should be regulated (Kanchiswamy et al., 2015). Current regulatory frameworks consider both the end product (i.e. the crop variety) and the methods used to develop it, but the relative importance differs between legislations. Canada's legislation on genetically engineered crops is focused on regulating the end product and would not consider a genome edited crop variety with no transgenic material as a transgenic crop (Smyth and McHughen, 2008). Regulation in the United States focuses primarily on the end product, but takes the processes into account (McHughen and Smyth, 2008). This stance has led to uncertainty about how non-transgenic genome edited crops are to be regulated (Wolt et al., 2016). In 2016, USDA issued the first decision not to regulate a CRISPR/Cas-edited crop variety under GM-regulation (Waltz, 2016). A US federal mandatory GM-food labelling law passed in July 2016 excludes products of genome editing. The EU and Japan are yet to issue legal frameworks for the regulation of genome-edited crops (Ledford, 2016). This divergence in regulation of these novel technologies among highly developed nations has created uncertainty among regulatory authorities in low-income countries. Hence it is not clear yet how non-transgenic genome-edited crops will be regulated in these countries in the future.

In jurisdictions that treat non-transgenic genome edited crop varieties the same as conventionally bred varieties, CRISPR/Cas edited zero-ODAP grass pea varieties could be introduced much more quickly and cheaply than in jurisdictions where such varieties would be classified as GMOs requiring extensive tests on their ecological and human health impacts. If the same trait could be achieved by TILLING, this issue would be avoided, because the TILLING process does not involve transgenic methods. In addition, the licencing costs of the CRISPR/Cas technology for agricultural applications may be prohibitive for use in grass pea, a crop of little commercial interest at present. Hence TILLING would currently be the method of choice for producing low-/zero-ODAP genotypes as breeding material for new varieties, while the CRISPR/Cas system would be primarily a research tool. If mutations at multiple loci are necessary to achieve a zero-ODAP genotype (e.g. if several BAHD-ATs in

the genome are capable of synthesising β-L-ODAP), this could be achieved by stacking mutations identified through TILLING, but using CRISPR/Cas with multiple guide RNAs to target each locus may be the faster approach.

5.6 Is L-DAP an intermediate in the synthesis of β-L-ODAP in grass pea?

L-2,3-diaminopropionic acid is generally regarded as the final intermediate in the synthesis of β-L-ODAP. Evidence for this was provided by Malathi et al. (1967; 1968; 1970) who partially purified an enzyme from grass pea that catalyses the formation of beta-ODAP in the presence of L-DAP and oxalyl-CoA. The authors noted that no measurable levels L-DAP were found in grass pea tissues and concluded that this compound must be a short lived intermediate in the synthesis. Subsequent analyses have similarly failed to demonstrate the presence of L-DAP in any grass pea tissue. My own analysis by means of LCMS using an external standard did not show detectable levels of L-DAP in any analysed grass pea sample, despite a very high level of sensitivity, as described in Chapter 3. In addition, none of the mutants in the entire mutant screen showed accumulation of L-DAP to an extent where it could be distinguished from the background using the spectrophotometric method. Although this screen primarily aimed to identify variations in ODAP content, the spectrophotometric assay that I was using relies on a colour-forming reaction between L-DAP, β-mercaptoethanol and o-phthalaldehyde. To measure β-L-ODAP content in a sample, the β-L-ODAP first has to be converted into L-DAP by alkaline hydrolysis. This posed a problem as mutants in which the last step of the synthesis was compromised might be accumulating L-DAP instead of β-L-ODAP, which cannot be distinguished after hydrolysis. To circumvent this issue, I tested each sample twice, using the hydrolysed and non-hydrolysed extract respectively. The absorbance value produced by the non-hydrolysed extract provides a measure for the L-DAP content of the sample. However, in the entire mutant screen, I observed no individuals with measurable levels of L-DAP. None of the low-ODAP mutants show measurable levels of L-DAP either. If this compound were an intermediate in the synthesis of β-L-ODAP, it would seem likely that some mutations might lead to it being over-accumulated.

Only one study has described evidence for the enzymatic formation of L-DAP (Ikegami et al., 1999). The authors describe the extraction of an enzyme from grass pea that catalyses the formation of L-DAP from the earlier pathway intermediate β-isoxazolin-alanine (BIA).

However, the optimum reaction conditions they describe (37 °C at pH 9.0, followed by termination by acidification to pH 1.8) appear far from physiological and the authors do not report whether this reaction also occurred under physiological conditions. They note, however, that at under slightly more alkaline conditions (pH 10) BIA begins to spontaneously degrade to L-DAP even in the absence of enzyme. This raises doubt over whether this enzyme catalyses the reaction in planta.

L-DAP itself is a stable compound and the kinetics of the β-L-ODAP-forming reaction described by Malathi et al. showed that not all L-DAP was converted into β-L-ODAP, even if oxalyl-CoA was supplied in excess (Malathi et al., 1970). This makes it unlikely that the in vivo reaction in grass pea would have its equilibrium so far in the direction of β-L-ODAP that no L-DAP would be detectable. To test whether another compound may be the substrate for β-L-ODAP production in grass pea, in vitro experiments using the grass pea ODAP-synthase to test alternative substrates as described earlier must be undertaken. BAHD-acyltransferases typically catalyse freely reversible reactions (Ma et al., 2005; Luo et al., 2009). The equilibrium generated between L-DAP and β-L-ODAP in vitro could give an indication on whether it was likely that L-DAP might be a short-lived intermediate that does not accumulate to detectable levels in vivo. The ODAP-synthase could also be supplied with other potential substrates, such as β-L-malonyl-diaminopropionic acid (β-L-MDAP), which is a hypothetical intermediate in the β-L-ODAP biosynthesis pathway (Yan et al., 2006). If ODAP-synthase is able to accept β-L-MDAP as a substrate, L-DAP may be bypassed entirely.

5.7   The physiological role of β-L-ODAP

The most important scientific question that can be addressed through the use of the low-ODAP mutants developed in this project and potentially zero-ODAP genotypes that could be developed in the future concerns the physiological role of β-L-ODAP in grass pea. By comparing the performance of near-isogenic high- and low-/zero-ODAP genotypes under conditions of abiotic stress it will be possible to ascertain whether β-L-ODAP is involved in stress responses. This is particularly relevant for the environmental stress factors that grass pea shows high tolerance to, i.e. drought and flooding. Insect feeding studies, again involving near isogenic high- and low-/zero-ODAP genotypes, are necessary to determine whether β-L-ODAP acts as a toxin or feeding deterrent that effectively reduces insect feeding. For these experiments, it would be necessary to include both known insect herbivores of grass pea, such as pea aphids (*Acyrtosiphon pisum*) (Wale and Gedif), pod

borer (*Etiala jhinkinella*) (Pandey et al., 1995) or thrips (*Caliothrips indicus*) (Pandey et al., 1995), as well as common insect pest species that are present in target areas for grass pea cultivation, but are not known pests of grass pea, because it is possible that these species are effectively deterred by the presence of β-L-ODAP.

In addition to these experiments testing specific hypotheses under laboratory conditions, it will be vital to test new low-/zero-ODAP genotypes under agricultural conditions in the target countries, e.g. Ethiopia, India or Bangladesh. This may reveal unexpected effects of the loss of β-L-ODAP in agricultural settings.

5.8   Platform development for the rapid domestication of grass pea

As part of this project, several new resources for the improvement of grass pea have been developed or are currently in development. Once released for use by other researchers and breeders, these tools will allow modern scientific and breeding techniques to be applied to the study and crop improvement of grass pea.

The transcriptomes of 8-day-old seedling shoot and root tips, leaves and roots of 5-week-old plants and flowers, early pods and late pods from 2-month-old plants of the Indian grass pea variety LSWT11 have been sequenced for this project. This represents the first transcriptome dataset covering different developmental stages and tissues of grass pea. The sequence database is currently being held on servers at JIC for internal use, but will be made available to outside researchers through the publicly available sequence database held by NCBI and a sequence server operated by JIC.

To expand on this transcriptome dataset, another transcriptome sequencing project is currently underway at JIC in collaboration with the Earlham Institute (EI), Norwich, UK and National Institute for Agricultural Botany (NIAB), Cambridge, UK, aimed at measuring the transcriptomic responses of grass pea to drought stress. The European grass pea variety LS007 and the Indian variety Mahateora are being used for this experiment, along with *Pisum sativum* cv. Cameor and *Vicia faba* cv. Hedin for comparison. RNA was extracted from roots and shoots of the plants. This dataset will help us to understand the genetic and physiological mechanisms that mediate the exceptional drought tolerance of grass pea. The identification of specific genes or overall regulatory patterns that confer this tolerance

could inform breeding approaches to improve the yield stability of other crops in the face of water stress.

As described in section 5.5, a grass pea draft genome is now being assembled. This draft genome will allow the identification of candidate gene sequences by automatic annotation of known patterns and through comparative genomics. These could serve as target genes for reverse genetic approaches, e.g. using TILLING or genome editing. The draft genome can be improved by the incorporation of further grass pea sequencing data to achieve better scaffolding and sequence coverage. This could be combined with genetic maps that are currently being developed (Wang et al., 2015; Yang et al., 2014) to produce high-quality physical maps that allow gene identification by map-based cloning.

As mentioned, a TILLING population based on the European variety LS007 is now being developed by the company BenchBio, Valvada, Gujarat, India. This population will serve as a tool for identifying mutant alleles in target genes. Hence, the platform could be used for the rapid improvement of grass pea by screening for mutant alleles in homologues of genes that have proved important in the breeding of other legumes. This could allow some of the genetic changes making up what has been called the 'domestication syndrome' (Weeden, 2007) to be established in grass pea through directed, reverse genetics methods. As an example, double podding has been an important trait for improving grain yields in legume crops (Hole and Hardwick, 1976). This trait has been observed in some grass pea accessions, but is not present in most cultivars (Campbell, 1997). Recently, the gene associated with double podding in chickpea has been identified (Ali et al., 2016). If a homologue to this gene exists in grass pea, it would be a promising target for developing the double podding trait in grass pea varieties.

The CRISPR/Cas platform that is now being developed could also be applied to develop agriculturally relevant traits other than low/zero-ODAP content. One possible application is the engineering of robust resistance to powdery mildew, caused by the fungus *Erysiphe pisi*. Powdery mildew is an agronomically significant disease affecting many crop species, including *Lathyrus sativus* (Vaz Patto et al., 2006). Broad-spectrum resistance has long been known in *Pisum sativum* (Heringa et al., 1969; Pavan et al., 2011) and a loss-of-function allele of the *MLO1* gene that confers resistance has recently been identified in this species (Sun et al., 2016a). Homologues of this gene family have been described in several other legume species (Rispail and Rubiales, 2016). Finding the homologue of the *PsMLO1* gene in

grass pea could provide a target for developing robust resistance to powdery mildew in grass pea by means of genome editing.

5.9 Bringing low-/zero-ODAP grass peas into the field

Before reliable low-ODAP or zero-ODAP grass pea varieties can be marketed as completely safe, it is crucial to prove that no risk of neurolathyrism remains when these varieties are used as food or feed. Because there still is no animal model that mirrors the symptoms of human neurolathyrism induced by consumption of grass pea, achieving this may be difficult. It would help the case for zero-ODAP grass peas if it could be shown that the acute neurotoxic effects that are observed in some animals after consumption of a diet heavily dependent on grass pea do not occur if zero-ODAP grass peas are used. However, the use of animal feeding studies raises ethical concerns, especially when using primate species, such as squirrel monkeys, which have previously been employed in β-L-ODAP feeding studies (Parker et al., 1979; Mehta et al., 1980, 1983). The improved understanding of the aetiology of neurolathyrism following more recent publications (Shinomol and Muralidhara, 2007; Van Moorhem et al., 2011; Nunn et al., 2011; Meiner and Gotkine, 2016) may help to identify an appropriate animal model species, as well as trackable symptoms of the disease, to minimise the number of animals required and the animal suffering caused by this research. To ensure the safety of grass pea varieties, regular measurements of β-L-ODAP levels in field-grown grass pea material should be made by food safety authorities.

If the problem of neurotoxicity can be resolved, the focus of the nutritional improvement of grass pea should shift to other characteristics, in particular antinutritional factors such as trypsin inhibitors and polyphenols and the seed protein amino acid profile, which is low in tryptophan, methionine and cysteine. In addition, *L. sativus* contains small amounts of 2-cyanoethyl-isoxazolin-5-one, which is catabolised to β-amino-propionitrile (BAPN), the causative agent of osteo- and angiolathyrism. Although this disease is primarily associated with *L. odoratus* and several wild species of *Lathyrus*, it may also be a health concern for the consumption of *L. sativus*. None of these problems represent barriers to application as major as neurotoxicity, but they still limit the value of grass pea as a major component of animal feed or a staple food for human consumption. The genetic and genomic resources that are now being developed by JIC and collaborating institutions will surely help in addressing these issues.

If the potential of grass pea for food security in regions prone to drought and flooding is to be realised, improved varieties need to be brought into cultivation both in countries that currently cultivate grass pea and countries that do not. Countries that already cultivate grass pea as a significant crop, such as Ethiopia, India and Bangladesh (Malek and Gazipur, 1999; Fikre et al., 2011; Bhowmick, 2013), may show greater interest in improved varieties because many farmers and consumers are already familiar with the crop. Grass pea is primarily cultivated by smallholder farmers who save seed from harvest to sowing or buy seed on the informal market (Campbell, 1997). This poses the risk of improved grass pea varieties hybridising with potentially unsafe landraces or even high-ODAP seed being intentionally mislabelled and sold as seed of improved varieties. Intentional mislabelling of non-improved seeds occurred when genetically modified Bt-cotton was first introduced in India; conventional cotton seed was sold in bags of the GM varieties by unscrupulous traders, leading to reports of the GM-crop failing (Stone, 2007; Sheridan, 2009).

One way to reduce the risk of high-ODAP grass pea landraces being mistaken for improved grass pea varieties and to identify events of unwanted outcrossing would be to generate improved varieties that also carry easily distinguishable morphological traits that set them apart from high-ODAP landraces and other varieties. This approach was already taken in the development of the low-ODAP variety Mahateora, which was bred to have pink flowers, instead of the more common blue flowers (Kumar et al., 2011a). Other traits that could be investigated for this purpose would be pod anthocyanin production, seed colour or seed shape. What traits are used to set improved varieties apart may need to depend on what traits are common in grass pea genotypes already cultivated in a region and what will be acceptable to consumers, who may be averse to unfamiliar traits, especially regarding seed morphology.

The high tolerance to environmental stress that grass pea exhibits makes it suitable to many areas where it is not currently cultivated or where its cultivation has ceased. In countries where grass pea has been used traditionally and is commonly associated with famine and disease, it may take time to overcome the image of grass pea as an undesirable food, even if improved varieties are advocated by government agencies or non-state actors such as non-governmental organisations or private companies. Areas where grass pea has never been introduced and areas where there has been historical cultivation of grass pea, but no living memory of neurolathyrism, may be good targets for the introduction of new toxin-free varieties, because the advantages of grass pea cultivation could be applied without having to overcome deeply held doubts over the safety of the crop. This may prove

easier to achieve by advocating grass pea as a green manure (i.e. a cover crop that is used to suppress weeds and regenerate the soil between plantings of other crops) or fodder crop, where consumer preference is less of a concern than for food crops. In addition, outcrossing with high-ODAP landraces would not be a concern in such areas. However, the fact that this is a new food unknown to most of the population may result in slow adoption of these varieties by food producers.

In either case, attempts to introduce improved grass pea varieties need to take account of existing seed systems in target countries. Presently, approximately 80 % of seed planted in Sub-Saharan Africa originates from informal seed systems, e.g. by farmers swapping, gifting or selling non-certified seed produced on-farm as well as saving seed from harvest to sowing (Byerlee et al., 2007). Functioning formal seed systems relying on breeding and release of certified crop varieties that are multiplied and distributed by the government or private seed companies exist, but primarily focus on a small number of crop species, such as maize or cotton (Smale et al., 2013). However, the prediction that formal seed systems would come to supplant the informal seed sector in Sub-Saharan Africa, as it has in developed countries, has not yet been fulfilled (Louwaars and De Boef, 2012; Lohr et al., 2015). Most pulse crops in Sub-Saharan Africa are primarily traded though the informal seed market (Almekinders et al., 1994). There is evidence that low-income farmers in Sub-Saharan Africa (Uganda, Burundi, Rwanda and Democratic Republic of Congo) are willing to buy improved bean seed if it provides a clear economic benefit (David and Sperling, 1999). Grass pea, however, is primarily grown as a food security crop with limited market demand at present. This may make it difficult to build an economically sustainable formal seed system for grass pea relying solely on private breeding companies selling seeds of improved varieties. Instead, an approach integrating publicly funded breeding of locally suitable genotypes (including a participatory approach to ensure the varieties meet the needs of local farmers under realistic farming conditions) with already existing informal seed systems may be more successful (Louwaars and De Boef, 2012; Louwaars et al., 2013; Lohr et al., 2015). This would require collaboration between international researchers and national agricultural research systems (such as EIAR in Ethiopia), government agricultural extension systems as well as national and/or international NGOs and farmer's collectives. This network of partners would be necessary to not only provide seeds (at no, or subsidised cost) to farmers, but also provide training on the benefits and limitations of these varieties, how to recognise improved varieties (using phenotypic markers, such as flower colour), and how to cultivate them successfully. In addition, this network could be utilised to allow

feedback of farmer's experiences to inform future varietal development. Networks like this are already in place in West Bengal, India, where researchers and breeders are engaging local farmers to multiply seed of certified low-ODAP varieties on-farm and provide training on their use (personal communication, Raghunath Sadhukhan, BCKV). The participatory breeding approach may also focus on the multiple possible uses of grass pea as a grain legume and a leafy vegetable, as a fodder crop and as a cover crop (Hillocks and Maruthi, 2012; Dixit et al., 2016) by developing cultivars specifically adapted to these different uses.

In general, grass pea production appears to be following patterns in production of other pulse crops. Ethiopia has seen rising acreages under grass pea cultivation since 2000 (Haimanot et al., 2005; Girma and Korbu, 2012). After declining between 1990 and 2009 (Miah and Haque, 2013), acreage under cultivation with grass pea in Bangladesh has remained level between 2009 and 2014, remaining the second most widely cultivated pulse crop, shortly after lentil (Bangladesh Bureau of Statistics, 2016). In the state of West Bengal, India, grass pea cultivation has doubled (after stagnating at a low level for six years) between 2014 and 2016, following increased demand for pulses and government-funded programs to distribute improved varieties (personal communication, Raghunath Sadhukhan, BCKV). At present, these data are too ambiguous to support either a falling or rising trend in global grass pea acreage.

In the longer term, the global demand for pulses is expected to increase as an effect of rising human and livestock populations and dietary changes. This increase is likely to be most pronounced in Southern Asia and Sub-Saharan Africa, due to the faster rate of population growth in these areas compared to global trends (Joshi and Rao, 2016). It has been projected that consumption of pulses in Sub-Saharan Africa will increase by 50 % between 2009 and 2030 (Clancey, 2009). New, locally adapted pulse varieties could contribute to alleviating the expected deficit in pulse production in South Asia and Sub-Saharan Africa (Clancey, 2009), reducing the need for imports and improving local food security. In the context of the impacts of climate change increasing water stress in these areas (Dai, 2013), drought tolerant crops such as grass pea could be instrumental in meeting this challenge.

Appendix

App. 1 Chapter 2 – Screening grass pea germplasm for low-ODAP genotypes

App. 1.1 Standard concentrations used alongside spectrophotometric assays

App. Table 1. Standard concentrations used alongside the single seed extracts comparing batches of Indian grass pea varieties and alongside the IPK germplasm population. Standards were produced by linear dilution of a L-DAP.HCl stock solution of 1.76 g/l. Equivalent seed ODAP wt% are given for 30 mg samples extracted with 600 µl buffer (single seed measurements) or 500 mg samples extracted with 10 ml buffer (IPK population)

% of L-DAP-HCl stock solution	0	0.25	0.5	1	2	3	4	5	6	7	8	10
L-DAP concentration (g/l)	0	0.004	0.009	0.018	0.035	0.053	0.070	0.088	0.106	0.123	0.141	0.176
Equivalent seed ODAP wt%	0	0.033	0.066	0.132	0.265	0.397	0.529	0.662	0.794	0.926	1.059	1.323

App. Table 2. Standard concentrations used alongside the USDA germplasm population. Standards were produced by linear dilution of a L-DAP.HCl stock solution of 2.0545 g/l. Equivalent seed ODAP wt% are given for 500 mg samples extracted with 10 ml buffer

% of L-DAP-HCl stock solution	0	0.5	1	2	4	6	8	10	12	14	16	20
L-DAP concentration (g/l)	0	0.010	0.021	0.041	0.082	0.123	0.164	0.205	0.247	0.288	0.329	0.411
Equivalent seed ODAP wt%	0	0.085	0.151	0.301	0.622	0.923	1.242	1.546	1.847	2.165	2.469	3.089

App. Table 3. Standards used alongside the EIAR germplasm population. Standards were produced by linear dilution of a L-DAP.HCl stock solution of 0.267 g/l. Equivalent seed ODAP wt% are given for 500 mg samples extracted with 10 ml buffer

% of L-DAP-HCl stock solution	0	10	20	30	40	50	60	70
L-DAP concentration (g/l)	0	0.027	0.053	0.080	0.106	0.133	0.160	0.186
Equivalent seed ODAP wt%	0	0.2	0.4	0.6	0.8	1	1.2	1.4

App. 1.2 Standard curves used for calibration of spectrophotometric assays

Standard curve used for comparing seed batches of Indian grass pea varieties
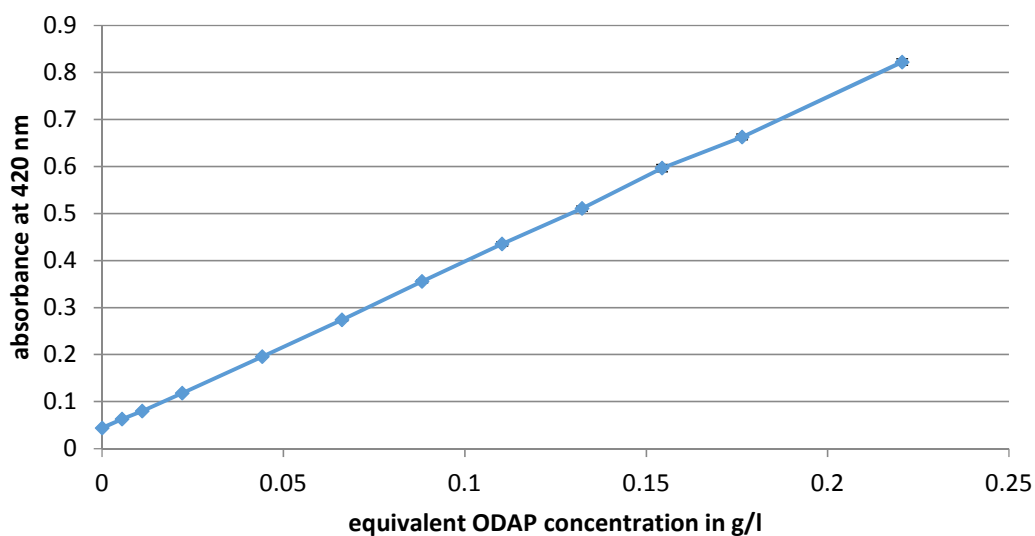
App. Figure 1. Calibration curve of L-DAP.HCl standards used with the assay on individual seeds of Indian grass pea varieties. Standards were included with each plate of samples. Error bars denote standard error. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	3.454 ± 0.013
intercept	0.042 ± 0.001
R^2	0.99986
degrees of freedom	10

Standard curve used for the IPK germplasm population
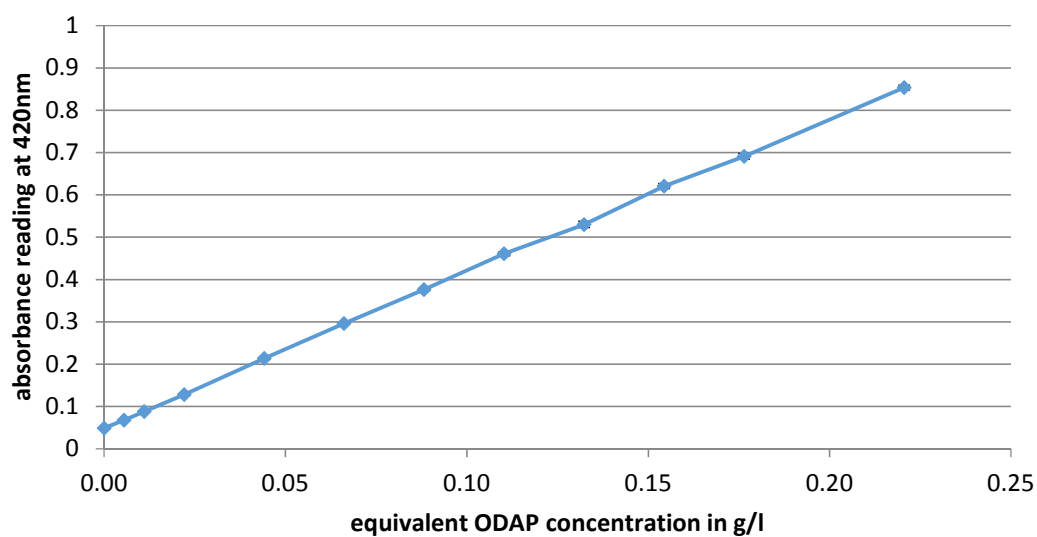
App. Figure 2. Calibration curve of L-DAP.HCl standards used with the IPK population. Standards were included with each plate of samples. Error bars denote standard error. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	3.659 ± 0.018
intercept	0.050 ± 0.002
R^2	0.99976
degrees of freedom	10

Standard curves used for the USDA population

The assays of the USDA population were performed in three sets on consecutive days. L-DAP standards were measured alongside each set of samples and separate standard curves were used to calculate seed ODAP concentrations.
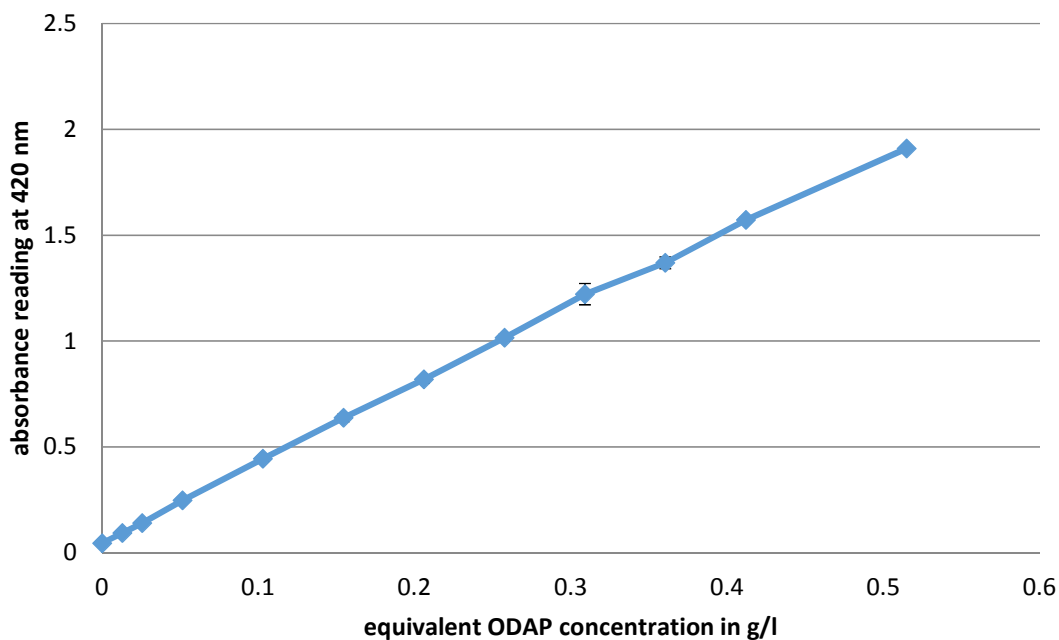
App. Figure 3. Calibration curve of L-DAP.HCl standards used with the first set of accessions (IDs 68-124) of the USDA population. Standards were included with each plate of samples. Error bars denote standard error of six replicates. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	3.659 ± 0.059
intercept	0.032 ± 0.008
R^2	0.99926
degrees of freedom	10
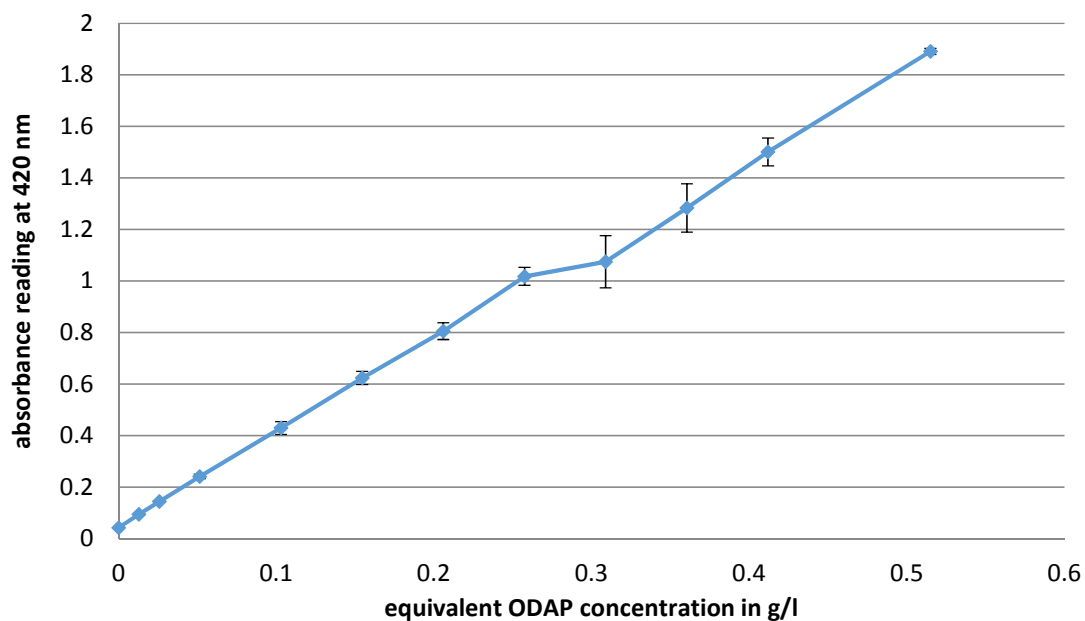
App. Figure 4. Calibration curve of L-DAP.HCl standards used with the second set of accessions (IDs 125-193) of the USDA population. Standards were included with each plate of samples. Error bars denote standard error of six replicates. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	3.510 ± 0.059
intercept	0.060 ± 0.016
R^2	0.99709
degrees of freedom	10
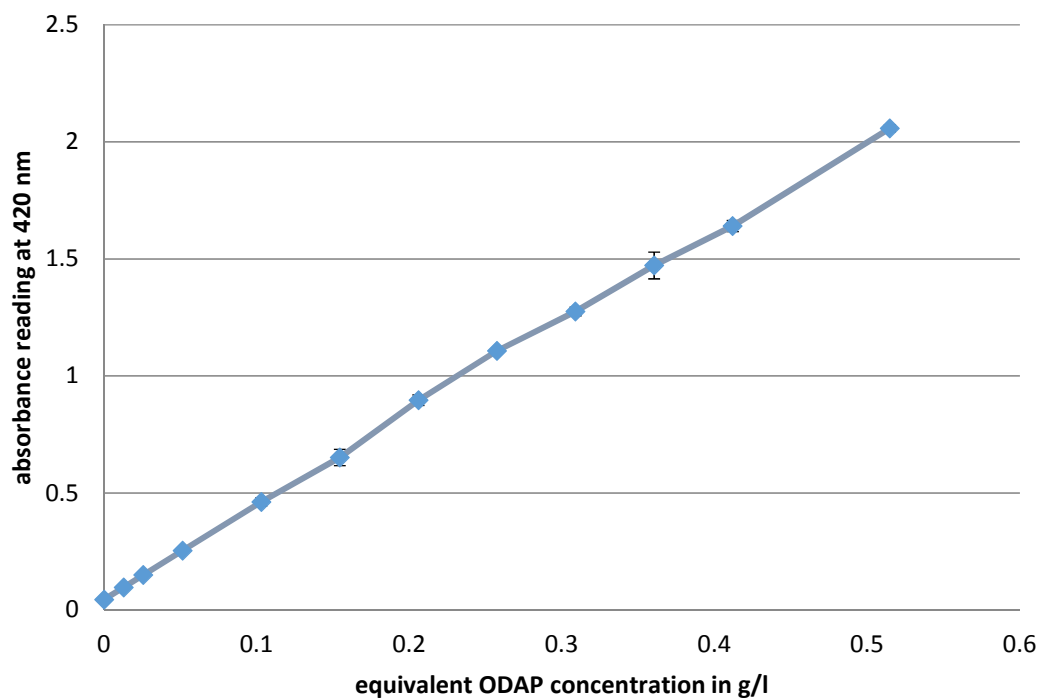
App. Figure 5. Calibration curve of L-DAP.HCl standards used with the third set of accessions (IDs 194-219) of the USDA population and the Indian grass pea varieties. Standards were included with each plate of samples. Error bars denote standard error of six replicates. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	3.916 ± 0.057
intercept	0.037 ± 0.010
R^2	0.99910
degrees of freedom	10

Standard curve used for the EIAR population
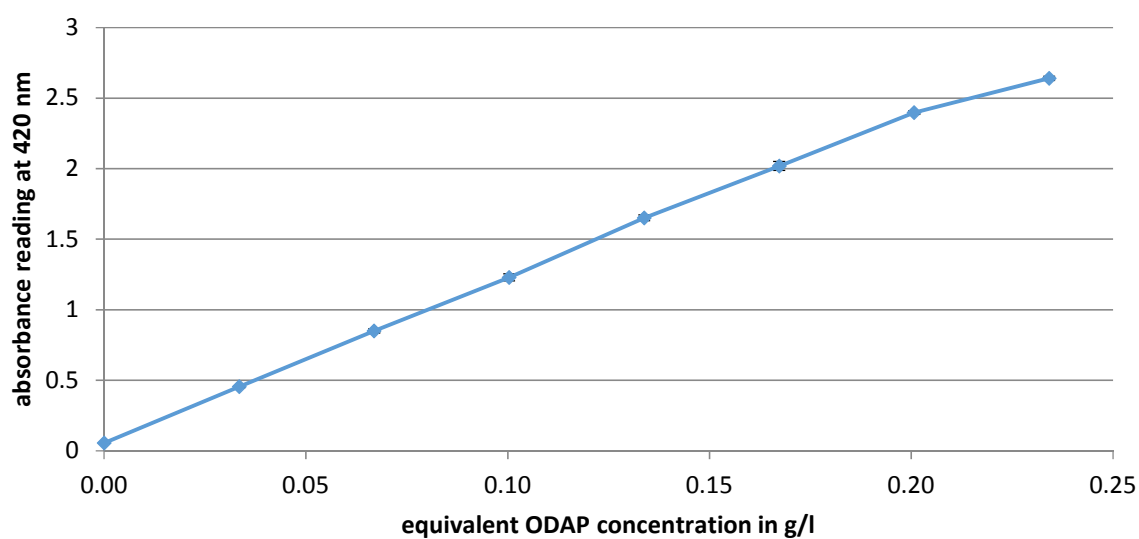
App. Figure 6. Calibration curve of L-DAP.HCl standards used with the EIAR population. Standards were included with each plate of samples. Error bars denote standard error. The table below shows the statistics of the linear regression used to calculate sample ODAP concentrations.

slope	11.296 ± 0.238
intercept	0.090 ± 0.033
R^2	0.99735
degrees of freedom	6

App. 1.3 Raw data of seed ODAP concentrations of grass pea accessions

IPK population

App. Table 4. Seed ODAP concentrations and countries of origin of the grass pea accessions obtained from the IPK collection, based on three technical replicate measurements of bulk seed samples. Empty cells denote incomplete passport data.

IPK ID	Seed ODAP concentration in % w/w	Country of origin
LAT 401	0.262	Turkey
LAT 404	0.308	Turkey
LAT 407	0.345	Greece
LAT 411	0.400	Greece
LAT 413	0.263	Greece
LAT 416	0.368	Greece
LAT 420	0.316	
LAT 422	0.351	
LAT 424	0.291	
LAT 430	0.259	Greece
LAT 431	0.347	Greece
LAT 432	0.286	Greece
LAT 435	0.311	Ukraine
LAT 436	0.245	
LAT 437	0.264	
LAT 446	0.333	
LAT 450	0.392	Iran
LAT 451	0.285	Iran
LAT 453	0.402	Tunisia
LAT 455	0.366	Hungary
LAT 463	0.321	Bulgaria
LAT 466	0.297	Russia
LAT 475	0.392	Slovakia
LAT 476	0.350	Slovakia
LAT 478	0.440	Czech Republic
LAT 494	0.301	Hungary
LAT 495	0.324	Slovakia
LAT 4003	0.298	Slovakia
LAT 4007	0.309	Spain
LAT 4008	0.337	Ethiopia
LAT 4010	0.277	Ethiopia
LAT 4011	0.250	Hungary

LAT 4024	0.267	Italy
LAT 4025	0.282	Italy
LAT 4032	0.330	Italy
LAT 4031	0.339	Italy
LAT 4036	0.216	Italy
LAT 4038	0.248	Italy
LAT 4039	0.252	Peru
LAT 4046	0.291	Italy
LAT 4047	0.277	Italy
LAT 4059	0.360	Slovakia
LAT 4075	0.321	Italy
LAT 4141	0.335	Italy

USDA population and Indian varieties

App. Table 5. Seed ODAP concentrations and countries of origin of the grass pea accessions obtained from the USDA collection and Indian varieties from BCKV, based on three technical replicate measurements of bulk seed samples. Empty cells denote incomplete passport data.

ID	USDA ID	Seed ODAP concentration in % w/w	Country of origin	Other names
68	PI 337087/A	0.509	Brazil	616
69	PI 337087/B	0.536	Brazil	616
70	PI 345525/A	0.307	India	Rewa-2
71	PI 345525/B	0.567	India	Rewa-2
72	PI 358857	0.466	Turkey	979
73	PI 358891	0.547	USA	48290
74	PI 366129	0.337	USA	China Pea
75	PI 370600	0.446	Former Serbia and Montenegro	Lokalen
76	PI 422528/A	0.289	Hungary	L-04
78	PI 422535/A	0.371	Turkey	L-5
79	PI 422535/B	0.384	Turkey	L-5
80	PI 422536/A	0.355	Italy	L-7
81	PI 422536/B	0.212	Italy	L-7
82	PI 422540/A	0.385	Italy	L-13
83	PI 422540/B	0.331	Italy	L-13
84	PI 422540/C	0.287	Italy	L-13
85	PI 422543	0.373	Hungary	
86	PI 429368	0.457	Iran	96
87	PI 506418/A	0.040	Canada	LS8246
88	PI 506418/B	0.345	Canada	LS8246
89	PI 507931/A	0.353	Hungary	140021
90	PI 507931/B	0.424	Hungary	140021
91	PI 511770	0.435	Peru	Alverge
92	PI 513244	0.701	Pakistan	Matri
93	PI 577139	0.484	Bulgaria	B92-103
94	PI 577141	0.491	Nepal	2423
95	PI 667238/A	0.621	Greece	falra bean
96	PI 667238/B	0.578	Greece	falra bean
97	PI 667239	0.659	Turkey	WKT 61
98	PI 667248	0.465	Bulgaria	Stranja
99	PI 667250/A	0.456	Albania	AI143
100	PI 667250/B	0.445	Albania	AI143
101	PI 667263	0.353	Georgia	9097
102	PI 667254	0.579	Tajikistan	ICG 137000

103	PI 667264/A	0.285	USA	IFLS170 Sel 439
104	PI 667264/B	0.447	USA	IFLS170 Sel 439
105	PI 667265/A	0.419	USA	IFLS273 Sel 481
106	PI 667265/B	0.452	USA	IFLS273 Sel 481
107	PI 667266/A	0.164	USA	IFLS 385 Sel 504
109	PI 667267	0.480	USA	IFLS 404 Sel 508
110	PI 667268/A	0.219	USA	IFLS 420 Sel 516
111	PI 667268/B	0.598	USA	IFLS 420 Sel 516
112	PI 667269/A	0.660	USA	IFLS 432 Sel 519
113	PI 667269/B	0.763	USA	IFLS 432 Sel 519
114	PI 667270/A	0.572	USA	IFLS 433 Sel 520
115	PI 667270/B	0.661	USA	IFLS 433 Sel 520
116	PI 667271	0.650	USA	IFLS 450 Sel 522
117	PI 667272	0.560	USA	IFLS 462 Sel 527
118	PI 667273/A	0.137	USA	IFLS 394 Sel 528
120	PI 667273/B	0.375	USA	IFLS 394 Sel 528
121	PI 667274/A	0.568	USA	IFLS 486 Sel 531
122	PI 667274/B	0.303	USA	IFLS 486 Sel 531
123	PI 667275/A	0.474	USA	IFLS 223 Sel 553
124	PI 667275/B	0.394	USA	IFLS 223 Sel 553
119	PI 667276/A	0.353	USA	IFLS 225 Sel 554
125	PI 667276/B	0.348	USA	IFLS 225 Sel 554
126	PI 667277	0.400	USA	IFLS 340 Sel 563
127	PI 667278/A	0.233	USA	IFLS 347 Sel 587
128	PI 667278/B	0.334	USA	IFLS 347 Sel 587
129	W6 9389	0.397	Pakistan	Wild Pea # 2
130	W6 25211/A	0.427	Tajikistan	ICC 136912
131	W6 25211/B	0.424	Tajikistan	ICC 136912
132	W6 28025	0.390	Tajikistan	TJK 2006:281
133	W6 39220/A	0.662	Bangladesh	Jamalpur
134	W6 39220/B	0.239	Bangladesh	Jamalpur
135	W6 39221	0.374	Ethiopia	Debre Zeit
136	W6 39222	0.244	Poland	Derek
137	W6 39225	0.226	Canada	LS87124
139	W6 39227	0.414	Bangladesh	Mymensingh
140	PI 163293	0.494	India	Teora
142	PI 165528/B	0.331	India	Chateri
143	PI 170469	0.138	Turkey	Murdumuk
144	PI 170470	0.324	Turkey	Fasil
145	PI 170477/A	0.378	Turkey	
146	PI 170477/B	0.455	Turkey	
148	PI 182780/B	0.461	Turkey	

149	PI 194995	0.353	Ethiopia	
150	PI 206891/A	0.186	Turkey	
151	PI 206891/B	0.446	Turkey	
152	PI 206892/A	0.385	Turkey	
153	PI 206892/B	0.496	Turkey	
154	PI 209789/A	0.371	German	Geissener Bunte Platterbse
155	PI 209789/B	0.325	German	Geissener Bunte Platterbse
156	PI 218082	0.362	Pakistan	
157	PI 221463	0.421	Afghanistan	Pateque
158	PI 221465	0.397	Afghanistan	Kalol; Pateque
159	PI 221466	0.450	Afghanistan	Kalol
160	PI 221467	0.456	Afghanistan	Kalol
161	PI 223270	0.409	Afghanistan	
162	PI 239865	0.579	Iran	
163	PI 239866	0.358	Iran	
164	PI 239867	0.462	Iran	
165	PI 244756	0.373	Ethiopia	
166	PI 255368/A	0.373	Former Serbia and Montenegro	
167	PI 255368/B	0.373	Former Serbia and Montenegro	
168	PI 269921	0.467	Pakistan	
169	PI 283546/A	0.541	Egypt	CPI 9512
170	PI 283546/B	0.450	Egypt	CPI 9512
171	PI 283547/A	0.391	France	CPI 9668
172	PI 283547/B	0.401	France	CPI 9668
173	PI 283548/A	0.342	Cyprus	Favetta
174	PI 283548/B	0.302	Cyprus	Favetta
175	PI 283550/A	0.233	Former Soviet Union	CPI 10724
176	PI 283550/B	0.190	Former Soviet Union	CPI 10724
177	PI 283552/A	0.315	Former Soviet Union	CPI 10726
178	PI 283552/B	0.243	Former Soviet Union	CPI 10726
179	PI 283553/A	0.346	Italy	CPI 10780
180	PI 283553/B	0.378	Italy	CPI 10780
181	PI 283557/A	0.452	Former Soviet Union	CPI 10786
182	PI 283557/B	0.460	Former Soviet Union	CPI 10786
183	PI 283559/A	0.299	Portugal	CPI 12411
184	PI 283559/B	0.360	Portugal	CPI 12411
185	PI 283560/A	0.237	Morocco	CPI 13977
186	PI 283560/B	0.247	Morocco	CPI 13977
187	PI 283561/A	0.253	Greece	CPI 14162
188	PI 283561/B	0.184	Greece	CPI 14162
189	PI 283562/A	0.351	India	CPI 14630
190	PI 283562/B	0.316	India	CPI 14630
191	PI 283562/C	0.469	India	CPI 14630
192	PI 283563/A	0.209	Spain	CPI 15232

193	PI 283563/B	0.244	Spain	CPI 15232
194	PI 283564	0.318	Sudan	CPI 15438
195	PI 283565/A	0.349	Morocco	CPI 15801
196	PI 283565/B	0.293	Morocco	CPI 15801
197	PI 283566/A	0.450	Morocco	CPI 15802
198	PI 283566/B	0.291	Morocco	CPI 15802
199	PI 283569/A	0.333	Libya	CPI 18401
200	PI 283569/B	0.291	Libya	CPI 18401
201	PI 283570/A	0.305	Algeria	Egypt
202	PI 283570/B	0.307	Algeria	Egypt
203	PI 283570/C	0.352	Algeria	Egypt
204	PI 283588/A	0.285	Czechoslovakia	CPI 22833
205	PI 283588/B	0.265	Czechoslovakia	CPI 22833
206	PI 283597/A	0.296	Tunisia	CPI 25091
207	PI 283597/B	0.426	Tunisia	CPI 25091
208	PI 283598/A	0.422	Tunisia	CPI 25092
209	PI 283598/B	0.254	Tunisia	CPI 25092
210	PI 283599/A	0.220	Tunisia	CPI 25093
211	PI 283599/B	0.251	Tunisia	CPI 25093
213	PI 286531	0.335	India	BN 13641-62
214	PI 317438/A	0.412	Afghanistan	113
215	PI 317438/B	0.440	Afghanistan	113
216	PI 317441	0.631	Afghanistan	Mashing
217	PI 317442	0.449	Afghanistan	546
218	PI 337007/A	0.398	Brazil	600
219	PI 337007/B	0.560	Brazil	600
P-24		0.153	India	
Mahateora		0.039	India	
Nirmal		0.167	India	
LS007		0.292	UK	
LSWT11		0.317	India	
Pea	*Pisum sativum*	0.001	UK	Cameor

EIAR population

App. Table 6. Seed ODAP concentrations and geographic origins of accessions comprising the EIAR population, based on three technical replicate measurements of bulk seed samples. Empty cells denote incomplete passport data.

Acc. Number	Region	Zone	Woreda/ District	Latitude	Longitude	Altitude	Seed ODAP % w/w
46104	Amhara	South Wello	Werebabu			1930	0.315
46107	Amhara	North Gondar	Gondar Zuria	12-37-00-N	37-10-00-E	1950	0.336
46108	Amhara	North Gondar	Gondar Zuria	12-37-00-N	37-10-00-E	2000	0.443
46109	Amhara	North Gondar	Gondar Zuria			2000	0.389
46110	Oromia	West Wellega	Nole Kaba	09-06-00-N	35-24-00-E	1600	0.284
46111	Amhara	South Wello	Tehuledere				0.316
201508	Oromia	East Shewa	Ada'a Chukala				0.332
201509	Oromia	East Shewa	Ada'a Chukala				0.278
201510	Oromia	East Shewa	Ada'a Chukala				0.385
201511	Oromia	East Shewa	Ada'a Chukala				0.308
201512	Oromia	East Shewa	Ada'a Chukala				0.286
201513	Oromia	East Shewa	Ada'a Chukala				0.296
201514	Oromia	East Shewa	Ada'a Chukala				0.381
201515	Oromia	East Shewa	Ada'a Chukala				0.398
201516	Oromia	East Shewa	Ada'a Chukala				0.310
201517	Oromia	East Shewa	Ada'a Chukala				0.338
201518	Oromia	East Shewa	Ada'a Chukala				0.342
201519	Oromia	East Shewa	Ada'a Chukala				0.324
201520	Oromia	East Shewa	Ada'a Chukala				0.332
201521	Oromia	East Shewa	Ada'a Chukala				0.409
201522	Oromia	East Shewa	Ada'a Chukala				0.329
201523	Oromia	East Shewa	Ada'a Chukala				0.342
201524	Oromia	East Shewa	Ada'a Chukala				0.344
201525	Oromia	East Shewa	Ada'a Chukala				0.348
201526	Oromia	East Shewa	Ada'a Chukala				0.392
201527	Oromia	East Shewa	Ada'a Chukala				0.365
201528	Oromia	East Shewa	Ada'a Chukala				0.372
201529	Oromia	East Shewa	Ada'a Chukala				0.338
201530	Oromia	East Shewa	Ada'a Chukala				0.358
201531	Oromia	East Shewa	Ada'a Chukala				0.307
201532	Oromia	East Shewa	Ada'a Chukala				0.319
201533	Oromia	East Shewa	Ada'a Chukala				0.314
201534	Oromia	East Shewa	Ada'a Chukala				0.404
201535	Oromia	East Shewa	Ada'a Chukala				0.502

201536	Oromia	East Shewa	Ada'a Chukala				0.312
201537	Oromia	East Shewa	Ada'a Chukala				0.305
201538	Oromia	East Shewa	Ada'a Chukala				0.317
201539	Oromia	East Shewa	Ada'a Chukala				0.368
201540	Oromia	East Shewa	Ada'a Chukala				0.402
201541	Oromia	East Shewa	Ada'a Chukala				0.352
201542	Oromia	East Shewa	Ada'a Chukala				0.331
201543	Oromia	East Shewa	Ada'a Chukala				0.306
201544	Oromia	East Shewa	Ada'a Chukala				0.358
201545	Oromia	East Shewa	Ada'a Chukala				0.388
201546	Oromia	East Shewa	Ada'a Chukala				0.286
201547	Oromia	East Shewa	Ada'a Chukala				0.289
201548	Oromia	East Shewa	Ada'a Chukala				0.325
201549	Oromia	East Shewa	Ada'a Chukala				0.348
201550	Oromia	East Shewa	Ada'a Chukala				0.361
201551	Oromia	East Shewa	Ada'a Chukala				0.320
201552	Oromia	East Shewa	Ada'a Chukala				0.313
201553	Oromia	East Shewa	Ada'a Chukala				0.373
201554	Oromia	North Shewa	Gerar Jarso				0.324
201555	Oromia	North Shewa	Gerar Jarso				0.336
201556	Oromia	North Shewa	Gerar Jarso				0.269
201557	Oromia	North Shewa	Gerar Jarso				0.373
201558	Oromia	North Shewa	Gerar Jarso				0.367
201559	Oromia	North Shewa	Gerar Jarso				0.375
201560	Oromia	North Shewa	Gerar Jarso				0.311
201561	Oromia	North Shewa	Gerar Jarso				0.317
201562	Oromia	North Shewa	Gerar Jarso				0.318
201563	Oromia	North Shewa	Gerar Jarso				0.387
201564	Oromia	North Shewa	Gerar Jarso				0.321
201565	Oromia	North Shewa	Gerar Jarso				0.329
201566	Oromia	North Shewa	Gerar Jarso				0.363
201567	Oromia	North Shewa	Gerar Jarso				0.366
201568	Oromia	North Shewa	Gerar Jarso				0.343
207565	Tigray	Misrakawi	Wukro				0.362
207566	Tigray	Debubawi	Hintalo Wajirat				0.367
207567	Tigray	Mehakelegnaw	Laelay Maychew				0.421
207991	Oromia	Illubabor	Bure	08-14-00-N	35-06-00-E		0.291
208449	Amhara	West Gojjam	Adet	11-35-00-N	37-17-00-E		0.356
208450	Amhara	South Gondar	Este	11-45-00-N	37-35-00-E		0.332
208451	Amhara	South Gondar	Fogera	11-50-00-N	37-35-00-E		0.389
208452	Amhara	South Gondar	Fogera				0.451
211511	Oromia	West Wellega	Nejo	09-34-00-N	35-22-00-E	1740	0.242
212740	Amhara	East Gojjam	Shebel Berenta	38-23-00-N	10-30-00-E	2410	0.392
212741	Amhara	South Gondar	Kemekem	37-42-00-N	12-06-00-E	2000	0.458

212742	Amhara	South Gondar	Kemekem	37-43-00-N	12-00-00-E	1880	0.423
213088	Amhara	South Wello	Were Ilu				0.328
213089	Amhara	South Wello	Were Ilu				0.363
213251	Tigray	Misrakawi	Wukro				0.307
213252	Tigray	Mehakelegnaw	Laelay Maychew				0.302
213253	Tigray	Debubawi	Endamehoni				0.309
214798	Amhara	North Wello	Dawuntna Delant				0.312
214799	Amhara	North Wello	Dawuntna Delant				0.327
214800	Amhara	South Wello	Debresina				0.364
214801	Amhara	South Wello	Kelala				0.340
214802	Amhara	South Wello	Sayint				0.320
214803	Amhara	South Wello	Kalu				0.328
215246	Amhara	South Wello	Tehuledere	11-23-00-N	39-38-00-E	1870	0.280
215247	Amhara	North Wello	Guba Lafto	11-50-00-N	39-31-00-E		0.405
215313	Amhara	East Gojjam	Hulet Ej Enese				0.350
215705	Amhara	South Wello	Kalu	11-04-00-N	39-44-00-E		0.294
215706	Amhara	South Wello	Ambasel	11-18-00-N	39-42-00-E	1950	0.323
219945	Tigray	Mehakelegnaw	Adwa	14-08-00-N	38-49-00-E	1870	0.446
219946	Tigray	Mehakelegnaw	Laelay Maychew	14-07-00-N	38-35-00-E	2080	0.302
219947	Tigray	Mehakelegnaw	Laelay Maychew	14-05-00-N	38-44-00-E	2120	0.369
219948	Tigray	Mehakelegnaw	Laelay Maychew	14-06-00-N	38-48-00-E	2140	0.358
219949	Tigray	Mehakelegnaw	Adwa	14-07-00-N	38-45-00-E	2150	0.432
219950	Tigray	Mehakelegnaw	Adwa	14-10-00-N	38-56-00-E	2230	0.378
219951	Tigray	Mehakelegnaw	Adwa				0.376
219952	Tigray	Mehakelegnaw	Naeder Adet	14-02-00-N	38-42-00-E	1700	0.365
220117	Eritrea	Maekel	Galanefi	15-17-00-N	38-53-00-E	2320	0.378
220118	Eritrea			14-43-00-N	38-52-00-E	1980	0.447
220119	Eritrea			14-51-00-N	38-49-00-E	1920	0.425
221717	Tigray	Debubawi	Enderta				0.395
221718	Tigray	Debubawi	Enderta	13-29-00-N	39-30-00-E	2150	0.425
223219	Tigray	Misrakawi	Wukro	14-10-00-N	39-33-00-E	1930	0.326
226001	Amhara	South Wello	Debresina	10-36-00-N	38-45-00-E	2400	0.404
226002	Amhara	South Wello	Debresina	10-33-00-N	38-44-00-E	2420	0.352
226003	Amhara	South Wello	Debresina	10-47-00-N	38-42-00-E	2485	0.306
226004	Amhara	South Wello	Debresina	10-47-00-N	38-39-00-E	2450	0.437
226005	Amhara	South Wello	Legambo				0.356
226006	Amhara	South Wello	Kelala	10-39-00-N	38-50-00-E	2500	0.336
226007	Amhara	South Wello	Kelala	10-37-00-N	38-48-00-E	2450	0.339
226008	Amhara	South Wello	Debresina	10-35-00-N	38-46-00-E	2400	0.350
226009	Amhara	South Wello	Legambo	10-46-00-N	38-53-00-E	2610	0.334
226010	Amhara	South Wello	Legambo	10-47-00-N	38-56-00-E	2640	0.320
226011	Amhara	South Gondar	Este	11-03-00-N	38-09-00-E	2520	0.307
226012	Amhara	South Gondar	Este	11-25-00-N	38-15-00-E	2500	0.322
226013	Amhara	South Gondar	Este	11-33-00-N	38-02-00-E	2485	0.317

226014	Amhara	South Gondar	Este	11-27-00-N	37-59-00-E	2645	0.346
226015	Amhara	South Gondar	Este	11-25-00-N	37-58-00-E	2380	0.342
226016	Amhara	South Gondar	Este	11-24-00-N	37-56-00-E	2345	0.325
226017	Amhara	North Gondar	Dembia	12-30-00-N	37-24-00-E	1905	0.456
226018	Amhara	North Gondar	Gondar Zuria	12-31-00-N	37-20-00-E	1990	0.413
226019	Amhara	Bahir Dar Special	Bahir Dar	11-32-00-N	37-26-00-E	1685	0.408
227196	Amhara	South Gondar	Ebenat				0.366
227197	Amhara	South Gondar	Ebenat				0.400
228495	Amhara	East Gojjam	Awabel				0.361
228496	Amhara	East Gojjam	Hulet Ej Enese				0.343
228497	Amhara	East Gojjam	Hulet Ej Enese				0.366
228498	Amhara	South Gondar	Este				0.349
228719	Oromia	West Shewa	Dendi				0.375
229173	Amhara	West Shewa	Lay Betna Tach Bet				0.357
229174	Amhara	West Shewa	Lay Betna Tach Bet				0.448
229175	Amhara	West Shewa	Weremo Wajetuna Mid				0.373
230011	Oromia	Arssi	Sherka	07-35-00-N	39-32-00-E	2400	0.416
230012	Oromia	Bale	Adaba	07-02-00-N	39-26-00-E	2450	0.309
230013	Oromia	Bale	Nensebo	07-02-00-N	39-31-00-E	2570	0.380
230014	Oromia	Bale	Ginir	07-08-00-N	40-41-00-E	1950	0.332
231325	Oromia	Arssi	Merti	08-35-00-N	39-52-00-E	1740	0.363
231326	Oromia	Arssi	Jeju				0.334
232281							0.490
232282							0.448
232283							0.427
232284							0.369
232285							0.492
233340							0.342
233341							0.479
233342							0.395
233657							0.392
233658							0.417
233659							0.394
233660							0.475
233661							0.430
233662							0.428
233798							0.415
233799							0.450
233800							0.458
233801							0.404
233802							0.414
233803							0.459
233804							0.388

234035	Tigray	Misrakawi	Tselemti	14-08-00-N	38-51-00-E	2060	0.321
234036	Tigray	Mehakelegnaw	Adwa	14-12-00-N	38-48-00-E	2100	0.329
234037	Tigray	Mehakelegnaw	Laelay Maychew	14-09-00-N	38-46-00-E	2100	0.389
234038	Tigray	Misrakawi	Ganta Afeshum	14-12-00-N	38-37-00-E	2150	0.378
234039	Tigray	Mehakelegnaw	Adwa	14-11-00-N	38-47-00-E	2150	0.391
234040	Tigray	Mehakelegnaw	Laelay Maychew	14-11-00-N	38-46-00-E	2120	0.284
234041	Tigray	Mehakelegnaw	Naeder Adet	14-04-00-N	38-43-00-E	2120	0.412
234042	Tigray	Mehakelegnaw	Naeder Adet	14-04-00-N	38-43-00-E	2100	0.357
234043	Tigray	Mehakelegnaw	Naeder Adet	14-04-00-N	38-43-00-E	2100	0.284
234044	Tigray	Mehakelegnaw	Naeder Adet	14-04-00-N	38-04-00-E	2100	0.358
234045	Tigray	Mehakelegnaw	Laelay Maychew	14-04-00-N	38-46-00-E	2130	0.399
234046	Tigray	Mehakelegnaw	Naeder Adet	13-54-00-N	37-44-00-E	2150	0.342
235018	Amhara	North Wello	Guba Lafto	11-43-00-N	39-31-00-E	1900	0.341
235123	Tigray	Debubawi	Hintalo Wajirat	12-58-00-N	39-34-00-E	1880	0.365
235124	Tigray	Debubawi	Hintalo Wajirat	12-59-00-N	39-32-00-E	1990	0.282
235125	Tigray	Misrakawi	Wukro	13-33-00-N	39-28-00-E	2180	0.441
235126	Tigray	Mehakelegnaw	Degua Temben	13-38-00-N	39-14-00-E	2250	0.300
235127	Tigray	Misrakawi	Wukro	13-38-00-N	39-15-00-E	1780	0.304
235128	Tigray	Debubawi	Enderta	13-41-00-N	39-16-00-E	1820	0.431
235129	Tigray	Debubawi	Enderta	13-09-00-N	39-16-00-E	1740	0.306
235130	Tigray	Debubawi	Mekele				0.387
235131	Tigray	Debubawi	Hintalo Wajirat	13-25-00-N	39-23-00-E	2150	0.352
236255	Amhara	South Wello	Kutaber				0.343
236256	Amhara	South Wello	Werebabu			2100	0.423
236543	Oromia	East Shewa	Akaki			2001	0.390
236544	Oromia	East Shewa	Ada'a Chukala			1920	0.352
236545	Oromia	East Shewa	Ada'a Chukala			1890	0.313
236546	Oromia	East Shewa	Ada'a Chukala			1900	0.340
236547	Oromia	East Shewa	Ada'a Chukala			2060	0.384
236548	Oromia	East Shewa	Ada'a Chukala			2270	0.354
236549	Oromia	East Shewa	Ada'a Chukala				0.359
236551	Oromia	East Shewa	Ada'a Chukala			2400	0.408
236552	Oromia	East Shewa	Ada'a Chukala			2440	0.368
236553	Oromia	East Shewa	Ada'a Chukala			2660	0.347
236554	Oromia	North Shewa	Berehna Aleltu			2480	0.364
236555	Oromia	North Shewa	Berehna Aleltu			2500	0.315
236557	Oromia	East Shewa	Ada'a Chukala			1890	0.297
236558	Oromia	East Shewa	Ada'a Chukala			1890	0.438
236559	Oromia	East Shewa	Ada'a Chukala			1890	0.357
236560	Oromia	East Shewa	Ada'a Chukala			1910	0.345
236561	Oromia	East Shewa	Ada'a Chukala			1890	0.482
236562	Oromia	East Shewa	Lome			1980	0.350
236563	Oromia	East Shewa	Lome			2020	0.378
236564	Oromia	East Shewa	Lome			2110	0.396

236565	Oromia	East Shewa	Lome				0.383
236566	Oromia	East Shewa	Lome			2160	0.318
236567	Oromia	East Shewa	Lome			2150	0.436
236568	Amhara	North Shewa	Minjarna Shenkora			2210	0.321
236569	Amhara	North Shewa	Minjarna Shenkora			2270	0.359
236570	Amhara	North Shewa	Minjarna Shenkora			2330	0.392
236571	Oromia	East Shewa	Gimbichu			2400	0.358
236572	Oromia	East Shewa	Gimbichu			2400	0.374
236573	Amhara	North Shewa	Minjarna Shenkora			2270	0.396
236574	Amhara	North Shewa	Minjarna Shenkora			2250	0.343
236575	Amhara	North Shewa	Minjarna Shenkora			2270	0.432
236576	Amhara	North Shewa	Minjarna Shenkora			2140	0.372
236577	Oromia	West Shewa	Alem Gena			2280	0.355
236578	Oromia	West Shewa	Walisona Goro			2080	0.358
236579	Oromia	West Shewa	Alem Gena			2050	0.406
236580	Oromia	West Shewa	Alem Gena			2050	0.402
236581	Oromia	West Shewa	Elu			2050	0.348
236582	Oromia	West Shewa	Elu			2050	0.344
236583	Oromia	West Shewa	Elu			2100	0.337
236584	Oromia	West Shewa	Becho			2140	0.326
236585	Oromia	West Shewa	Becho			2170	0.369
236586	Oromia	West Shewa	Becho			2230	0.294
236587	Oromia	West Shewa	Becho			2280	0.298
236588	Oromia	West Shewa	Becho			2320	0.420
236589	Oromia	West Shewa	Walisona Goro			2140	0.307
236590	Oromia	West Shewa	Walisona Goro			1900	0.439
236591	Oromia	West Shewa	Ambo			2220	0.363
236592	Oromia	West Shewa	Ambo			2140	0.342
236593	Oromia	West Shewa	Ambo			2320	0.287
236594	Oromia	West Shewa	Dendi			2410	0.371
236595	Oromia	West Shewa	Dendi			2400	0.342
236596	Oromia	West Shewa	Dendi			2280	0.372
236597	Oromia	West Shewa	Dendi			2280	0.367
236598	Oromia	West Shewa	Dendi			2240	0.366
236599	Oromia	West Shewa	Dendi			2190	0.318
236600	Oromia	West Shewa	Dendi			2107	0.350
236624	Oromia	West Shewa	Ambo			2240	0.365
236625	Oromia	North Shewa	Berehna Aleltu			2450	0.381
236626	Oromia	North Shewa	Berehna Aleltu			2450	0.373
236627	Oromia	North Shewa	Berehna Aleltu			2480	0.329
236628	Oromia	North Shewa	Berehna Aleltu			2580	0.398

236629	Amhara	North Shewa	Hageremaria-mina Kese				0.273
236630	Oromia	North Shewa	Kembibit			2810	0.355
236631	Oromia	West Shewa	Meta Robi			2850	0.302
236632	Amhara	North Shewa	Siyadebrina Wayu Ens			2630	0.366
236633	Amhara	North Shewa	Siyadebrina Wayu Ens			2670	0.307
236634	Amhara	North Shewa	Siyadebrina Wayu Ens			2630	0.320
236635	Amhara	North Shewa	Siyadebrina Wayu Ens			2600	0.325
236636	Amhara	North Shewa	Siyadebrina Wayu Ens			2620	0.333
236637	Amhara	North Shewa	Moretna Jiru			2640	0.295
236638	Amhara	North Shewa	Moretna Jiru			2650	0.306
236639	Amhara	North Shewa	Moretna Jiru			2640	0.353
236640	Amhara	North Shewa	Moretna Jiru			2640	0.375
236641	Amhara	North Shewa	Moretna Jiru			2630	0.316
236642	Amhara	North Shewa	Siyadebrina Wayu Ens			2650	0.359
236643	Amhara	North Shewa	Siyadebrina Wayu Ens			2660	0.413
236644	Amhara	North Shewa	Siyadebrina Wayu Ens			2620	0.455
236645	Amhara	North Shewa	Siyadebrina Wayu Ens				0.381
236646	Amhara	North Shewa	Siyadebrina Wayu Ens			2620	0.458
236647	Amhara	North Shewa	Siyadebrina Wayu Ens			2620	0.452
236648	Amhara	North Shewa	Siyadebrina Wayu Ens			2590	0.444
236649	Oromia	North Shewa	Wuchalena Jido				0.488
236650	Oromia	North Shewa	Wuchalena Jido			2660	0.504
236651	Oromia	North Shewa	Wuchalena Jido			2650	0.385
236652	Oromia	North Shewa	Yaya Gulelena D/Liba			2640	0.388
236653	Oromia	North Shewa	Yaya Gulelena D/Liba			2660	0.394
236654	Oromia	North Shewa	Yaya Gulelena D/Liba			2680	0.410
236655	Oromia	North Shewa	Gerar Jarso			2660	0.416
236656	Oromia	North Shewa	Gerar Jarso			2770	0.479
236657	Oromia	North Shewa	Gerar Jarso			2740	0.435
236658	Oromia	North Shewa	Gerar Jarso			2770	0.308
236659	Oromia	North Shewa	Gerar Jarso			2800	0.391
236660	Oromia	North Shewa	Gerar Jarso			2900	0.469
236661	Oromia	North Shewa	Kuyu			2580	0.465
236662	Oromia	North Shewa	Kuyu			2560	0.416
236663	Oromia	North Shewa	Wara Jarso			2560	0.367
236664	Oromia	North Shewa	Wara Jarso			2580	0.439
236665	Oromia	North Shewa	Wara Jarso			2520	0.404

236666	Amhara	East Gojjam	Dejen			2460	0.383
236667	Amhara	East Gojjam	Dejen			2460	0.405
236668	Amhara	East Gojjam	Dejen			2450	0.452
236669	Amhara	East Gojjam	Enemay			2430	0.388
236670	Amhara	East Gojjam	Enemay			2420	0.388
236671	Amhara	East Gojjam	Enemay			2450	0.469
236672	Amhara	East Gojjam	Enemay			2440	0.502
236673	Amhara	East Gojjam	Shebel Berenta			2420	0.417
236674	Amhara	East Gojjam	Shebel Berenta			2400	0.403
236675	Amhara	East Gojjam	Enemay			2420	0.393
236676	Amhara	East Gojjam	Enemay			2480	0.421
236677	Amhara	East Gojjam	Enemay			2428	0.421
236678	Amhara	East Gojjam	Enemay			2540	0.378
236679	Amhara	East Gojjam	Enemay			2470	0.397
236680	Amhara	East Gojjam	Enemay				0.437
236681	Amhara	East Gojjam	Enemay			2500	0.487
236682	Amhara	East Gojjam	Enarj Enawga				0.419
236683	Amhara	East Gojjam	Enarj Enawga			2560	0.376
236684	Amhara	East Gojjam	Enarj Enawga			2580	0.416
236685	Amhara	East Gojjam	Goncha Siso Enese			2580	0.453
236686	Amhara	East Gojjam	Goncha Siso Enese			2650	0.423
236687	Amhara	East Gojjam	Mota			2490	0.400
236688	Amhara	East Gojjam	Mota			2500	0.409
236689	Amhara	East Gojjam	Mota			2440	0.378
236690	Amhara	East Gojjam	Mota			2420	0.431
236691	Amhara	East Gojjam	Mota			2430	0.373
236692	Amhara	East Gojjam	Mota			2240	0.396
236693	Amhara	West Gojjam	Adet			1770	0.485
236694	Amhara	West Gojjam	Adet			2320	0.296
236695	Amhara	West Gojjam	Adet			2320	0.458
236696	Amhara	West Gojjam	Adet			2260	0.358
236697	Amhara	West Gojjam	Adet			2260	0.358
236698	Amhara	Bahir Dar Special	Bahir Dar			2230	0.419
236699	Amhara	Bahir Dar Special	Bahir Dar			2160	0.400
236700	Amhara	Bahir Dar Special	Bahir Dar			1830	0.380
236701	Amhara	Bahir Dar Special	Bahir Dar			1800	0.607
236702	Amhara	South Gondar	Dera			1800	0.455
236703	Amhara	South Gondar	Dera				0.538
236704	Amhara	South Gondar	Fogera			1800	0.505
236705	Amhara	South Gondar	Fogera			1800	0.496
236706	Amhara	South Gondar	Fogera			1800	0.447
236707	Amhara	South Gondar	Gondar Zuria			1920	0.387
236708	Amhara	South Gondar	Dabat			2730	0.382
236709	Amhara	Bahir Dar Special	Bahir Dar			1840	0.327

236710	Amhara	Bahir Dar Special	Bahir Dar			1880	0.392
236711	Amhara	Bahir Dar Special	Bahir Dar			1810	0.356
236712	Amhara	West Gojjam	Jabi Tehnan			1820	0.389
236713	Amhara	West Gojjam	Jabi Tehnan			1840	0.391
236714	Amhara	West Gojjam	Bahir Dar Zuria			1840	0.321
236715	Amhara	East Gojjam	Debre Markos			2300	0.370
236716	Amhara	East Gojjam	Awabel			2330	0.401
236717	Amhara	East Gojjam	Awabel			2410	0.439
236718	Amhara	East Gojjam	Awabel			2420	0.412
236719	Amhara	East Gojjam	Dejen			2680	0.325
236720	Oromia	North Shewa	Wuchalena Jido			2680	0.415
236721	Oromia	North Shewa	Wuchalena Jido			2660	0.347
236722	Oromia	North Shewa	Wuchalena Jido			2680	0.376
236993	Amhara	North Shewa	Minjarna Shenkora	08-50-00-N	39-20-00-E	1690	0.371
236994	Oromia	East Shewa	Ada'a Chukala	08-50-00-N	39-00-00-E	1820	0.340
237524	Tigray	Mehakelegnaw	Degua Temben	13-30-00-N	39-28-00-E	1770	0.331
237525	Tigray	Misrakawi	Wukro	13-46-00-N	39-35-00-E	2500	0.369
237526	Tigray	Mehakelegnaw	Laelay Maychew	14-10-00-N	38-45-00-E	2150	0.322
237527	Tigray	Mehakelegnaw	Tahtay Maychew	14-10-00-N	38-45-00-E	2100	0.390
237977	Oromia	Bale	Adaba			2440	0.287
238238	Tigray	Debubawi	Mekele	12-50-00-N	38-07-00-E	2000	0.310
238239	Tigray	Debubawi	Mekele	13-30-00-N	40-30-00-E	2250	0.338
238240	Tigray	Mehakelegnaw	Degua Temben	13-30-00-N	38-50-00-E	2560	0.381
238242	Tigray	Mehakelegnaw	Naeder Adet	13-40-00-N	38-50-00-E	2050	0.316
238243	Tigray	Mehakelegnaw	Laelay Maychew	13-40-00-N	38-50-00-E	2170	0.326
238244	Tigray	Mehakelegnaw	Laelay Maychew	14-50-00-N	39-60-00-E	2050	0.323
238245	Tigray	Misrakawi	Tahtay Koraro	16-60-00-S	38-50-00-E	1940	0.331
238901	Oromia	North Shewa	Mulona Sululta	09-23-85-N	09-23-85-E	2630	0.288
238902	Oromia	North Shewa	Mulona Sululta	09-27-08-N	38-52-02-E	2600	0.330
238903	Oromia	North Shewa	Yaya Gulelena D/Liba	09-36-07-N	38-51-07-E	2680	0.349
238904	Oromia	North Shewa	Wuchalena Jido	09-39-44-N	38-49-20-E	2680	0.300
238906	Oromia	North Shewa	Wara Jarso	09-53-22-N	38-21-77-E	2580	0.342
238908	Amhara	East Gojjam	Dejen	09-57-28-N	38-18-26-E	2450	0.345
238909	Amhara	East Gojjam	Dejen	10-13-80-N	38-07-79-E	2470	0.336
238910	Amhara	East Gojjam	Enemay	10-21-29-N	38-08-96-E	2440	0.363
238911	Amhara	East Gojjam	Enemay			2440	0.296
238914	Amhara	East Gojjam	Enarj Enawga	10-29-22-N	38-10-07-E	2520	0.311
238915	Amhara	East Gojjam	Enarj Enawga	10-43-11-N	38-09-57-E	2600	0.346
238917	Amhara	East Gojjam	Mota	10-44-35-N	38-09-37-E	2540	0.279
238919	Amhara	Bahir Dar Special	Bahir Dar	11-34-05-N	19-48-00-E	1890	0.368
238920	Amhara	West Gojjam	Merawi	11-24-69-N	37-09-54-E	2050	0.394
238921	Amhara	Bahir Dar Special	Bahir Dar	11-35-39-N	37-26-26-E	1900	0.327
238922	Amhara	South Gondar	Kemekem	11-51-03-N	37-38-69-E	1850	0.388
238923	Amhara	South Gondar	Kemekem	11-56-78-N	37-42-69-E	1860	0.380

238925	Amhara	South Gondar	Kemekem			1860	0.432
238926	Amhara	South Gondar	Kemekem	12-11-35-N	37-40-42-E	1950	0.459
238927	Amhara	North Gondar	Gondar Zuria			2000	0.354
238928	Amhara	North Gondar	Gondar Zuria	12-23-61-N	37-33-09-E	1990	0.357
238929	Amhara	North Gondar	Gondar Zuria			1960	0.429
238930	Amhara	North Gondar	Gondar	12-46-30-N	37-37-55-E	2110	0.447
238931	Amhara	South Gondar	Fogera	11-57-45-N	37-42-95-E	1920	0.355
238932	Amhara	South Gondar	Fogera	11-56-82-N	37-45-07-E	1920	0.247
238933	Amhara	South Gondar	Fogera	11-55-92-N	37-50-12-E	2040	0.331
238934	Amhara	South Gondar	Fogera	11-55-40-N	37-51-46-E	2080	0.414
238935	Amhara	South Gondar	Fogera	11-55-24-N	37-54-21-E	2130	0.392
238936	Amhara	Bahir Dar Special	Bahir Dar	11-29-72-N	37-31-66-E	1750	0.333
238937	Amhara	Bahir Dar Special	Bahir Dar	11-37-34-N	37-15-55-E	1940	0.403
238938	Amhara	Bahir Dar Special	Bahir Dar	11-40-04-N	37-12-43-E	1900	0.396
238939	Amhara	West Gojjam	Achefer	11-39-94-N	36-55-55-E	2070	0.375
238940	Amhara	West Gojjam	Achefer			2080	0.329
238941	Amhara	West Gojjam	Achefer	11-43-36-N	36-58-27-E	2000	0.314
238942	Amhara	West Gojjam	Achefer	11-44-33-N	36-59-06-E	2050	0.440
238943	Amhara	West Gojjam	Achefer	11-51-88-N	37-01-06-E	1910	0.406
238944	Amhara	West Gojjam	Achefer	11-50-00-E	36-59-97-E	2020	0.388
238945	Amhara	West Gojjam	Achefer	11-47-62-N	36-59-59-E	2030	0.384
238946	Amhara	West Gojjam	Dega Damot	10-44-66-N	37-19-68-E	1910	0.303
238947	Amhara	West Gojjam	Dega Damot	10-38-55-N	37-23-58-E	1900	0.347
238948	Amhara	East Gojjam	Guzamn	10-21-00-N	37-37-32-E	2340	0.369
238949	Amhara	East Gojjam	Awabel	10-16-16-N	37-49-43-E	2370	0.378
238950	Amhara	East Gojjam	Guzamn	10-14-31-N	37-57-18-E	2430	0.311
238951	Amhara	East Gojjam	Awabel	10-30-40-N	37-59-19-E	2460	0.379
238952	Amhara	East Gojjam	Awabel	10-94-40-N	35-03-72-E	2470	0.325
238953	Oromia	North Shewa	Wara Jarso			2550	0.304
238954	Oromia	North Shewa	Kuyu	10-00-56-N	38-14-94-E	2550	0.345
238956	Oromia	East Shewa	Akaki	08-49-98-N	38-50-25-E	2140	0.280
238960	Oromia	North Shewa	Berehna Aleltu			2450	0.320
238961	Oromia	North Shewa	Berehna Aleltu			2450	0.373
238962	Oromia	North Shewa	Berehna Aleltu	09-06-13-N	38-58-61-E	2470	0.322
238963	Oromia	North Shewa	Berehna Aleltu			2550	0.318
238965	Oromia	North Shewa	Berehna Aleltu	09-10-98-N	39-08-50-E	2620	0.341
238966	Oromia	North Shewa	Berehna Aleltu			2630	0.336
238967	Oromia	West Shewa	Meta Robi			2840	0.333
240031	Australia						0.327
240032	Australia						0.117
240033	Australia						0.092
240034	Australia						0.444
240035	Australia						0.035
240036	Australia						0.060

240037	Australia						0.372
240038	Australia						0.061
240039	Australia						0.114
240040	Australia						0.040
241143	Somali	Jigjiga	Jigjiga	09-30-18-N	42-37-12-E	1920	0.370
242216	Amhara	North Gondar	Gondar Zuria	12-30-00-N	37-32-00-E	1975	0.388
242217	Amhara	North Wello	Guba Lafto	11-50-00-N	39-35-00-E	1910	0.136
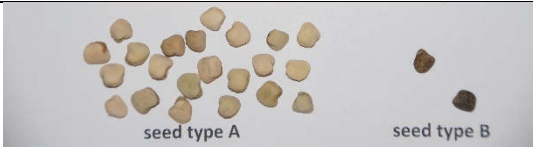
App. 1.4 Seed morphologies of seeds obtained from USDA

App. Table 7. List of accessions obtained from the USDA collection with photographs of the seeds delivered in each packet.

USDA ID	Country of origin	Other names	Photograph of seeds obtained from USDA (showing separation into sub-accessions)
PI 337087	Brazil	616	
PI 345525	India	Rewa-2	
PI 358857	Turkey	979	
PI 358891	USA	48290	
PI 366129	USA	China Pea	
PI 370600	Former Serbia and Montenegro	Lokalen	
PI 422528	Hungary	L-04	
PI 422535	Turkey	L-5	
PI 422536	Italy	L-7	
PI 422540	Italy	L-13	

PI 422543	Hungary		Uniform seed morphology
PI 429368	Iran	96	Uniform seed morphology
PI 506418	Canada	LS8246	seed type A seed type B
PI 507931	Hungary	140021	seed type A seed type B
PI 511770	Peru	Alverge	Uniform seed morphology
PI 513244	Pakistan	Matri	Uniform seed morphology
PI 577139	Bulgaria	B92-103	Uniform seed morphology
PI 577141	Nepal	2423	Uniform seed morphology
PI 667238	Greece	falra bean	seed type A seed type B
PI 667239	Turkey	WKT 61	Uniform seed morphology
PI 667248	Bulgaria	Stranja	Uniform seed morphology

PI 667250	Albania	AI143	seed type A seed type B
PI 667263	Georgia	9097	Uniform seed morphology
PI 667254	Tajikistan	ICG 137000	Uniform seed morphology
PI 667264	USA	IFLS170 Sel 439	seed type A seed type B
PI 667265	USA	IFLS273 Sel 481	seed type A seed type B
PI 667266	USA	IFLS 385 Sel 504	seed type A seed type B
PI 667267	USA	IFLS 404 Sel 508	Uniform seed morphology
PI 667268	USA	IFLS 420 Sel 516	seed type A seed type B
PI 667269	USA	IFLS 432 Sel 519	seed type A seed type B
PI 667270	USA	IFLS 433 Sel 520	seed type A seed type B
PI 667271	USA	IFLS 450 Sel 522	Uniform seed morphology

PI 667272	USA	IFLS 462 Sel 527	Uniform seed morphology
PI 667273	USA	IFLS 394 Sel 528	seed type A — seed type B
PI 667274	USA	IFLS 486 Sel 531	seed type A — seed type B
PI 667275	USA	IFLS 223 Sel 553	seed type A — seed type B
PI 667276	USA	IFLS 225 Sel 554	seed type A — seed type B
PI 667277	USA	IFLS 340 Sel 563	Uniform seed morphology
PI 667278	USA	IFLS 347 Sel 587	seed type A — seed type B
W6 9389	Pakistan	Wild Pea # 2	Uniform seed morphology
W6 25211	Tajikistan	ICC 136912	seed type A — seed type B
W6 28025	Tajikistan	TJK 2006:281	Uniform seed morphology
W6 39220	Bangladesh	Jamalpur	seed type A — seed type B

W6 39221	Ethiopia	Debre Zeit	Uniform seed morphology
W6 39222	Poland	Derek	Uniform seed morphology
W6 39225	Canada	LS87124	Uniform seed morphology
W6 39227	Bangladesh	Mymensingh	Uniform seed morphology
PI 163293	India	Teora	Uniform seed morphology
PI 165528/B	India	Chateri	seed type A seed type B
PI 170469	Turkey	Murdumuk	Uniform seed morphology
PI 170470	Turkey	Fasil	Uniform seed morphology
PI 170477	Turkey		seed type A seed type B
PI 182780	Turkey		seed type A seed type B
PI 194995	Ethiopia		Uniform seed morphology
PI 206891	Turkey		seed type A seed type B

PI 206892	Turkey		
PI 209789	Germany	Geissener Bunte Platterbse	
PI 218082	Pakistan		
PI 221463	Afghanistan	Pateque	
PI 221465	Afghanistan	Kalol; Pateque	
PI 221466	Afghanistan	Kalol	
PI 221467	Afghanistan	Kalol	
PI 223270	Afghanistan		
PI 239865	Iran		
PI 239866	Iran		
PI 239867	Iran		
PI 244756	Ethiopia		

PI 255368	Former Serbia and Montenegro		 seed type A seed type B
PI 269921	Pakistan		 Uniform seed morphology
PI 283546	Egypt	CPI 9512	 seed type A seed type B
PI 283547	France	CPI 9668	 seed type A seed type B
PI 283548	Cyprus	Favetta	 seed type A seed type B
PI 283550	Former Soviet Union	CPI 10724	 seed type A seed type B
PI 283552	Former Soviet Union	CPI 10726	 seed type A seed type B
PI 283553	Italy	CPI 10780	 seed type A seed type B
PI 283557	Former Soviet Union	CPI 10786	 seed type A seed type B
PI 283559	Portugal	CPI 12411	 seed type A seed type B
PI 283560	Morocco	CPI 13977	 seed type A seed type B
PI 283561	Greece	CPI 14162	 seed type A seed type B
PI 283562	India	CPI 14630	 seed type A seed type B seed type C

PI 283563	Spain	CPI 15232	seed type A · seed type B
PI 283564	Sudan	CPI 15438	Uniform seed morphology
PI 283565	Morocco	CPI 15801	seed type A · seed type B
PI 283566	Morocco	CPI 15802	seed type A · seed type B
PI 283569	Libya	CPI 18401	seed type A · seed type B
PI 283570	Algeria	Egypt	seed type A · seed type B · seed type C
PI 283588	Czechoslovakia	CPI 22833	seed type A · seed type B
PI 283597	Tunisia	CPI 25091	seed type A · seed type B
PI 283598	Tunisia	CPI 25092	seed type A · seed type B
PI 283599	Tunisia	CPI 25093	seed type A · seed type B · seed type C
PI 286531	India	BN 13641-62	Uniform seed morphology
PI 317438	Afghanistan	113	seed type A · seed type B
PI 317441	Afghanistan	Mashing	Uniform seed morphology

| PI 317442 | Afghanistan | 546 | |
| PI 337007 | Brazil | 600 | |

App. 1.5 Segregation in seed morphologies within sub-accessions
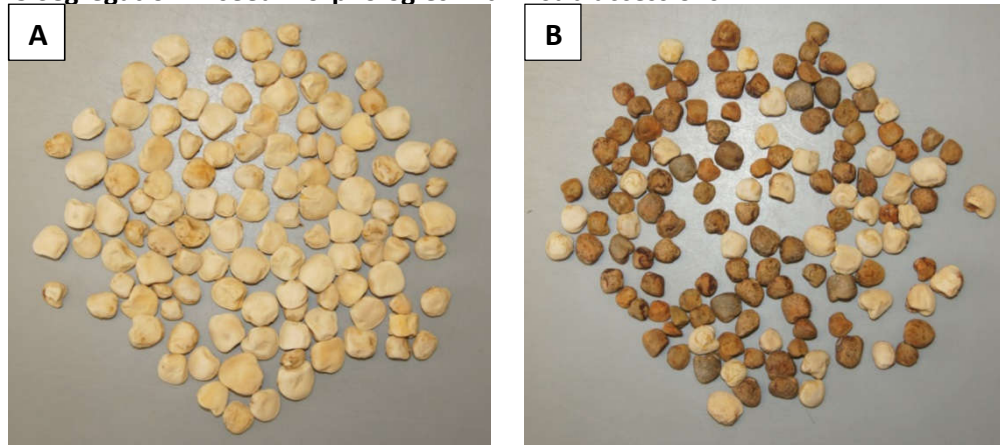
App. Figure 7. Seeds produced by sub-accessions of PI 255368 from the USDA population A) The seed type of sub-accession A was reproduced in the next generation B) the seeds produced by sub-accession B segregated for seed morphology in a roughly 3:1 ratio

App. 2 Chapter 3 – Identification of low-ODAP grass peas from a mutagenised population

App. 2.1 Example metadata file

An example of the plate metadata files, which were saved in .txt format for each plate of collected samples is shown below. An html-form was used to generate these files. The files contain information about which columns from which growth tray were harvested into the plate, dates of sowing, harvesting and assaying, initials of the person performing these procedures and which samples were missing because the seed in the relevant position on the growth tray had not germinated. These metadata were used by the data handling script to match absorbance readings to mutant families and to exclude missing samples.

```
plate ID        P13
origin tray T6  2,3,4,5,6,7,8,9
date sowed      28/04/14
sowed by        KK
date harvested  05/05/14
harvested by    KK

family miss 1  2  3  4  5  6  7  8  9  10 11 12
 1120  A
 1119  B                                    1
 1117  C
 1116  D
 1113  E              1                      1
 1112  F
 1109  G        1                               1
 1105  H                          1
date assayed
assayed by  KK
Notes       09/05/14
```

App. 2.2 R-script for selection of low-ODAP and high-background samples during the mutant screen

```
# helper function to extract plate IDs from the file names of metadata files
extractIDs<-function(vector){
  return(paste(vector[4:(length(vector)-4)],collapse=""))
}

# helper function to make a vector of the columns on a tray that were
# harvested to make one plate
getTrayColumns<-function(cols){
  splitcols<-strsplit(as.character(cols),split=",")
```

```
  return(as.numeric(splitcols[[1]]))
}
```

```
# main function. The parameters give the thresholds according to which samples
# will get selected for manual curation. Samples were marked as low-ODAP if their
# estimated ODAP content was less than ODAPthresh times the median value of their
# plate. Samples were marked as high-background if their non-hydrolysed reading
# was higher than backthresh. The outputdata argument is primarily for debugging,
# normally the data is saved directly into a .csv file that can be opened in other
# applications
```

```
ODAPscreen<-function(ODAPthresh=0.5,backthresh=0.09,outputdata=FALSE){
  setwd("U:/ODAP mutant screen/metadata")
  metaFiles<-list.files(pattern="*.txt",full.names=TRUE)
  splitnames<-strsplit(metaFiles,split="")
  plateVec<-sort(as.numeric(sapply(X=splitnames,FUN=extractIDs)))

  setwd("U:/ODAP mutant screen/plate reads")
  nonhydFiles<-list.files(pattern="*N.txt",full.names=TRUE)
  hydFiles<-list.files(pattern="*H.txt",full.names=TRUE)
  Nplates<-length(metaFiles)
  if(length(nonhydFiles)!=length(hydFiles)|length(nonhydFiles)!=Nplates) {
    stop("Data error: numbers of files have to be equal. Found ", Nplates," metafiles,
",length(hydFiles)," hydrated and ",length(nonhydFiles)," non-hydrated files")
  }
#data structure of data as they are being read in from the metadata and absorbance
#reading files:
  #plates[[i]][[1]]: plate ID
  #plates[[i]][[1]]: tray(s) of origin
  #plates[[i]][[2]][[1]]: origin tray 1
  #plates[[i]][[2]][[2]]: columns from origin tray 1
  #plates[[i]][[2]][[3]]: origin tray 2 (if applicable)
  #plates[[i]][[2]][[4]]: columns from origin tray 2 (if applicable)
  #plates[[i]][[2]][[5]]: origin tray 3 (if applicable)
  #plates[[i]][[2]][[6]]: columns from origin tray 3 (if applicable)
  #plates[[i]][[2]][[7]]: origin tray 4 (if applicable)
  #plates[[i]][[2]][[8]]: columns from origin tray 4 (if applicable)
  #plates[[i]][[3]]: dates and people
  #plates[[i]][[3]][[1]]: sowing date
  #plates[[i]][[3]][[2]]: sowed by
  #plates[[i]][[3]][[3]]: harvesting date
  #plates[[i]][[3]][[4]]: harvested by
  #plates[[i]][[3]][[5]]: assay date
  #plates[[i]][[3]][[6]]: assayed by
  #plates[[i]][[4]]: M2 families
  #plates[[i]][[5]]: missing samples matrix
  #plates[[i]][[6]]: non-hydrolysed OPT-absorbance data matrix
  #plates[[i]][[7]]: hydrolysed OPT-absorbance data matrix
  #plates[[i]][[8]]: approximate ODAP levels in samples (no normalisation by weight) in % of
dry weight matrix
  #plates[[i]][[9]]: notes
```

```r
plates<-list()

allS<-data.frame(        sampleID=numeric(0),
Line=character(0),
Tray=character(0),
Traypos=character(0),
Plate=character(0),
Platepos=character(0),
Missing=numeric(0),
NonHyd=numeric(0),
Hyd=numeric(0),
ODAP=numeric(0),
Keep=numeric(0),
stringsAsFactors=FALSE)

for (i in plateVec){
  setwd("U:/ODAP mutant screen/metadata")
  platemeta<-
read.csv(paste("./P",i,".txt",sep=""),header=FALSE,sep="\t",stringsAsFactors=FALSE,
col.names=c(1:14))
  setwd("U:/ODAP mutant screen/plate reads")
  nonPlate<-
read.csv(paste("./P",i,"N.txt",sep=""),sep="\t",header=FALSE,stringsAsFactors=FALSE)
  hydPlate<-
read.csv(paste("./P",i,"H.txt",sep=""),sep="\t",header=FALSE,stringsAsFactors=FALSE)
  setwd("U:/ODAP mutant screen")
  plates[[i]]<-list()
  plates[[i]][[1]]<-platemeta[1,2]
  plates[[i]][[2]]<-list()
  plates[[i]][[2]][[1]]<-platemeta[2,2]
  plates[[i]][[2]][[2]]<-getTrayColumns(platemeta[2,3])
# second tray. This code is run more than once because some plates contain samples
# from several growth trays, but never more than four
  if (!(is.na(platemeta[2,4]))){
    plates[[i]][[2]][[3]]<-platemeta[2,4]
    plates[[i]][[2]][[4]]<-getTrayColumns(platemeta[2,5])}
  else {plates[[i]][[2]][[4]]<-numeric()}
# third tray
  if (!(is.na(platemeta[2,6]))){
    plates[[i]][[2]][[5]]<-platemeta[2,6]
    plates[[i]][[2]][[6]]<-getTrayColumns(platemeta[2,7])}
  else {plates[[i]][[2]][[6]]<-numeric()}
# fourth tray
  if (!(is.na(platemeta[2,8]))){
    plates[[i]][[2]][[7]]<-platemeta[2,8]
    plates[[i]][[2]][[8]]<-getTrayColumns(platemeta[2,9])}
  else {plates[[i]][[2]][[8]]<-numeric()}

  plates[[i]][[3]]<-list()
  plates[[i]][[3]][[1]]<-platemeta[3,2]
  plates[[i]][[3]][[2]]<-platemeta[4,2]
```

```
   plates[[i]][[3]][[3]]<-platemeta[5,2]
   plates[[i]][[3]][[4]]<-platemeta[6,2]
   plates[[i]][[3]][[5]]<-platemeta[17,2]
   plates[[i]][[3]][[6]]<-platemeta[18,2]
   plates[[i]][[4]]<-platemeta[8:15,1]
   plates[[i]][[5]]<-matrix(       as.numeric(as.matrix(platemeta[8:15,3:14])),
ncol=12,nrow=8,dimnames=list(LETTERS[1:8],c(1:12)))
   plates[[i]][[6]]<-matrix(       as.numeric(as.matrix(nonPlate[4:11,3:14])),
ncol=12,nrow=8,dimnames=list(LETTERS[1:8],c(1:12)))
   plates[[i]][[7]]<-matrix(       as.numeric(as.matrix(hydPlate[4:11,3:14])),
ncol=12,nrow=8,dimnames=list(LETTERS[1:8],c(1:12)))
   plates[[i]][[8]]<-(plates[[i]][[7]]-plates[[i]][[6]])*0.17613/0.68/3*100
   plates[[i]][[9]]<-platemeta[2,19]

#data are reformatted for processing
   #data structure:
   #allS[s,1]: sampleID (running number)
   #allS[s,2]: Family (BenchBio family number)
   #allS[s,3]: Tray (Tray ID, e.g. T3)
   #allS[s,4]: Traypos (position on tray, letters are rows, numbers are columns, i.e. numbers
count along long side of tray)
   #allS[s,5]: Plate (Plate ID, e.g. P5)
   #allS[s,6]: Platepos (position on plate, letters are rows, numbers are columns, as marked
on plate)
   #allS[s,7]: Missing (1 if no sample was collected from this position)
   #allS[s,8]: NonHyd (raw absorbance reading of non-hydrated sample in OPT-assay)
   #allS[s,9]: Hyd (raw absorbance reading of hydrated sample in OPT-assay)
   #allS[s,10]: ODAP estimate (approximate dry weight ODAP level, assuming 3mg dry
sample (not weighed))
   #allS[s,11]: Keep (1 if this plant should be kept)


   allSlen<-(which(plateVec == i)-1)*96
   for (s in 1:96){
    platecol<-((s-1)%/%8)+1
    platerow<-s-(platecol-1)*8
    platepos<-paste(LETTERS[platerow],platecol,sep="")
    allS[s,1]<-allSlen+s
    allS[s,2]<-plates[[i]][[4]][platerow]

   if((9-platerow)<=length(plates[[i]][[2]][[2]])){
     traycol<-plates[[i]][[2]][[2]][length(plates[[i]][[2]][[2]])-(platerow-
length(plates[[i]][[2]][[4]])-length(plates[[i]][[2]][[6]])-length(plates[[i]][[2]][[8]]))+1]
     allS[s,3]<-plates[[i]][[2]][[1]]}
   else if((9-platerow)<=length(plates[[i]][[2]][[2]])+length(plates[[i]][[2]][[4]])){
     traycol<-plates[[i]][[2]][[4]][length(plates[[i]][[2]][[4]])-(platerow-
length(plates[[i]][[2]][[6]])-length(plates[[i]][[2]][[8]]))+1]
     allS[s,3]<-plates[[i]][[2]][[3]]}
   elseif((9platerow)<=length(plates[[i]][[2]][[2]])
                            +length(plates[[i]][[2]][[4]])+length(plates[[i]][[2]][[6]])){
```

```
      traycol<-plates[[i]][[2]][[6]][length(plates[[i]][[2]][[6]])-(platerow-
length(plates[[i]][[2]][[8]]))+1]
      allS[s,3]<-plates[[i]][[2]][[5]]}
    else {
      traycol<-plates[[i]][[2]][[8]][length(plates[[i]][[2]][[8]])-platerow+1]
      allS[s,3]<-plates[[i]][[2]][[7]]}

    traypos<-paste(LETTERS[platecol],traycol,sep="")
    # if(platepos %in% plates[[i]][[4]]){     # deprecated
    #   traypos<-paste("M",traycol,sep="")
    # }
    allS[s,4]<-traypos
    allS[s,5]<-paste("P",i,sep="")
    allS[s,6]<-platepos

    if (!is.na(plates[[i]][[5]][s]))  {allS[s,7]<-1}
    else    {allS[s,7]<-0}

    allS[s,8]<-plates[[i]][[6]][s]
    allS[s,9]<-plates[[i]][[7]][s]
    allS[s,10]<-plates[[i]][[8]][s]

    # selection of low-ODAP samples
    if((plates[[i]][[8]][s]<=(ODAPthresh*median(plates[[i]][[8]])) & allS[s,7]==0)){
      allS[s,11]<-1}
    else if(plates[[i]][[6]][s]>=(median(plates[[i]][[6]])+2*sd(plates[[i]][[6]])) & allS[s,7]==0){
      allS[s,11]<-2}
    else {allS[s,11]<-0}
  }
  cat("done with plate",i,"!\n")
  if(!(outputdata)){

write.table(allS,"screenResults.csv",col.names=FALSE,row.names=FALSE,na="",sep=",",app
end=TRUE,dec=".",qmethod="double")
    allS<-data.frame(stringsAsFactors=FALSE)
  }
 }
# data output into a .csv file
 if(outputdata){
 write.csv(allS,"screenResults.csv",row.names=FALSE,na="")
 return(allS)}
}
```

App. 2.3 LCMS calibration for single seed measurements (using external β-L-ODAP standards)
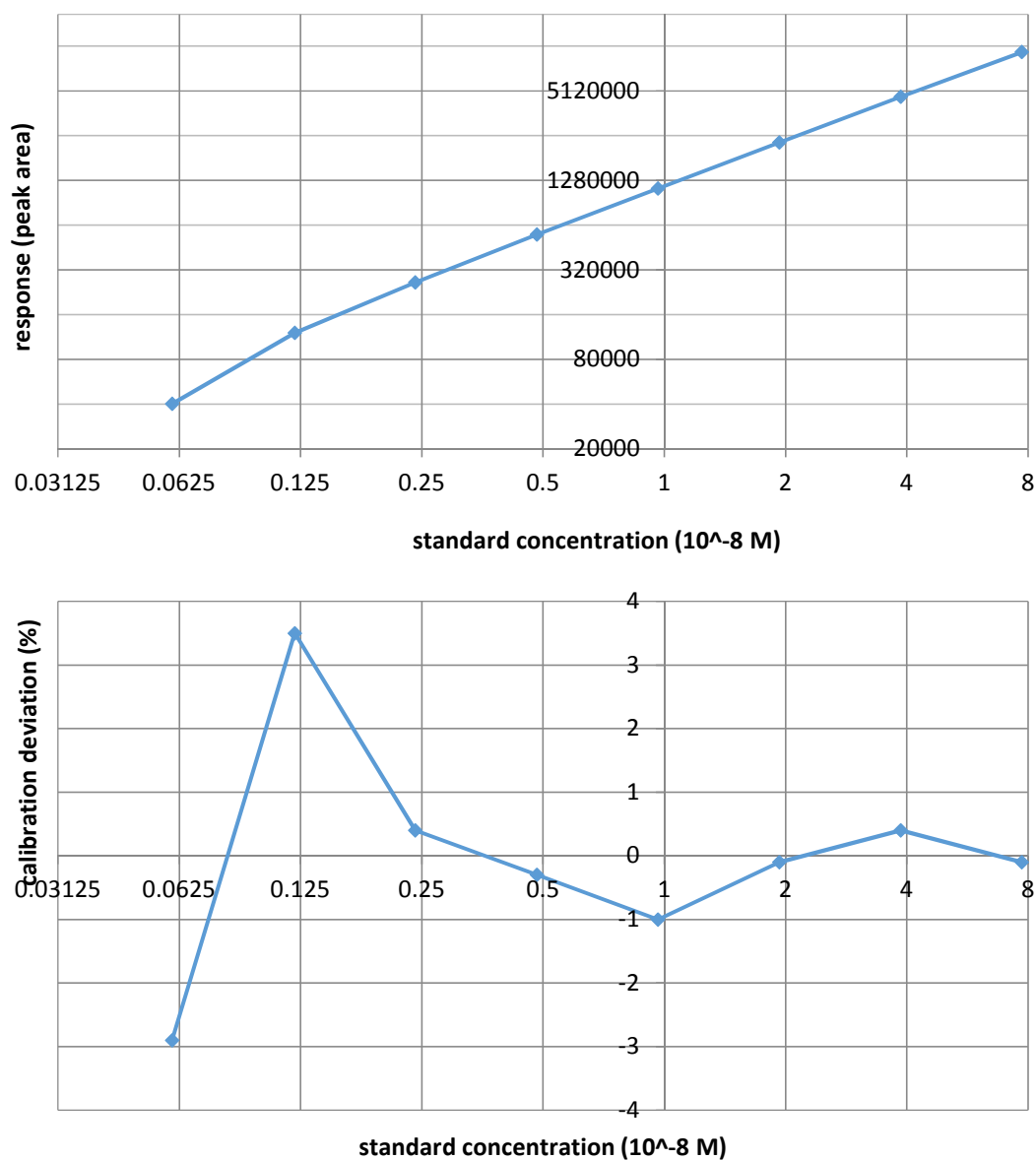
App. Figure 8. Calibration of LCMS using external β-L-ODAP standards for the measurement of single seed samples. A linear regression was calculated (see below) and used to calculate the ODAP concentrations of seed samples. All axes are logarithmic (base 2).

slope	$8.2348 * 10^{-7} \pm 8.8 * 10^{-10}$
intercept	0.02639 ± 0.00335
R^2	0.99999
degrees of freedom	6

App. 2.4 Calibration curves used to calculate β-L-ODAP concentrations of grass pea tissue samples measured by LCMS using the internal standard

Batch 1

Calibration curve	$-0.0207414\ x^2 + 5.49907\ x - 0.00807413$
R^2	0.999828
Samples in this batch, unless otherwise indicated, all three replicate extractions were in this batch	LSWT11 – leaf, flower, seed, root, early pod, late pod Mahateora – early pod, late pod, flower (reps 1,2) 1264-2 – seedling shoot tip, seed, seedling root tip (reps 1,2) Pea – seed
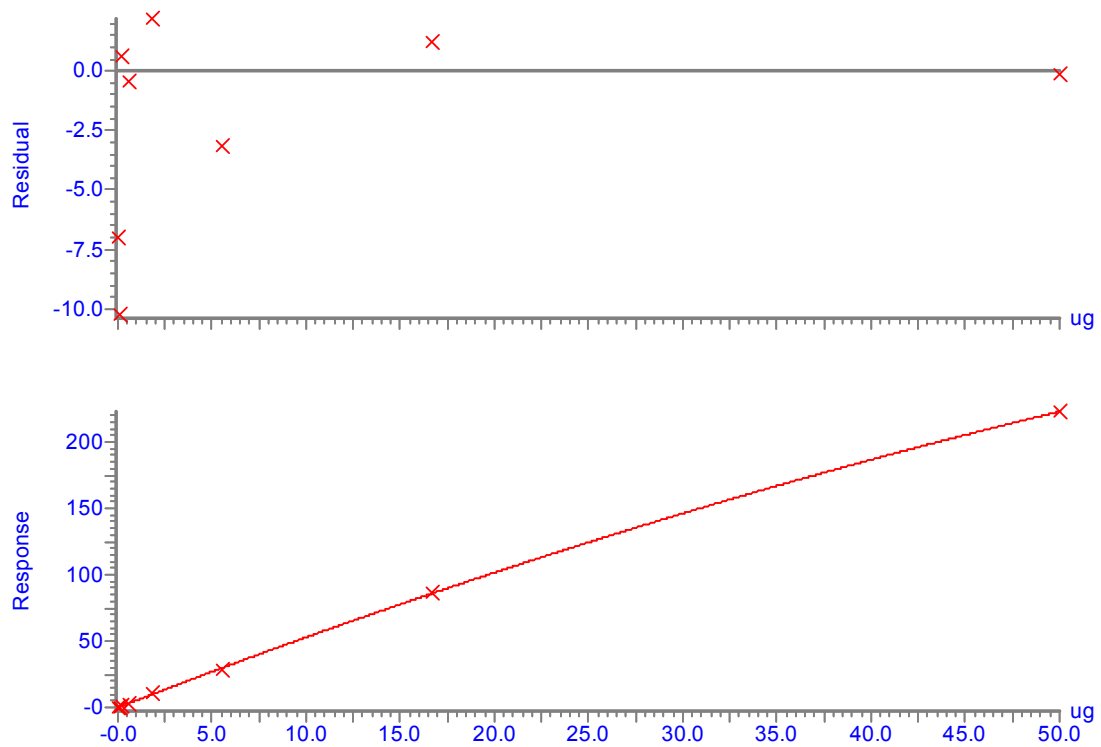
App. Figure 9. Calibration of LCMS using internal ^{13}C β-L-ODAP standards for the measurement of β-L-ODAP concentrations in batch 1 of the grass pea tissue samples

Batch 2

Calibration curve	$-0.0188358 \, x^2 + 4.74805 \, x - 0.052804$
R^2	0.999793
Samples in this batch, unless otherwise indicated, all three replicate extractions were in this batch	Mahateora – flower (rep 3), seed, seedling root tip, late pod, seedling shoot tip, leaf, root 1264-2 –seedling root tip (rep 3), flower, early pod, late pod, leaf, root 4884-2 – seedling root tip, seedling shoot tip, flower, seed, early pod, late pod, leaf, root 4946-7 – flower, root, seedling root tip, seed, early pod, late pod, leaf Pea – seedling root tip
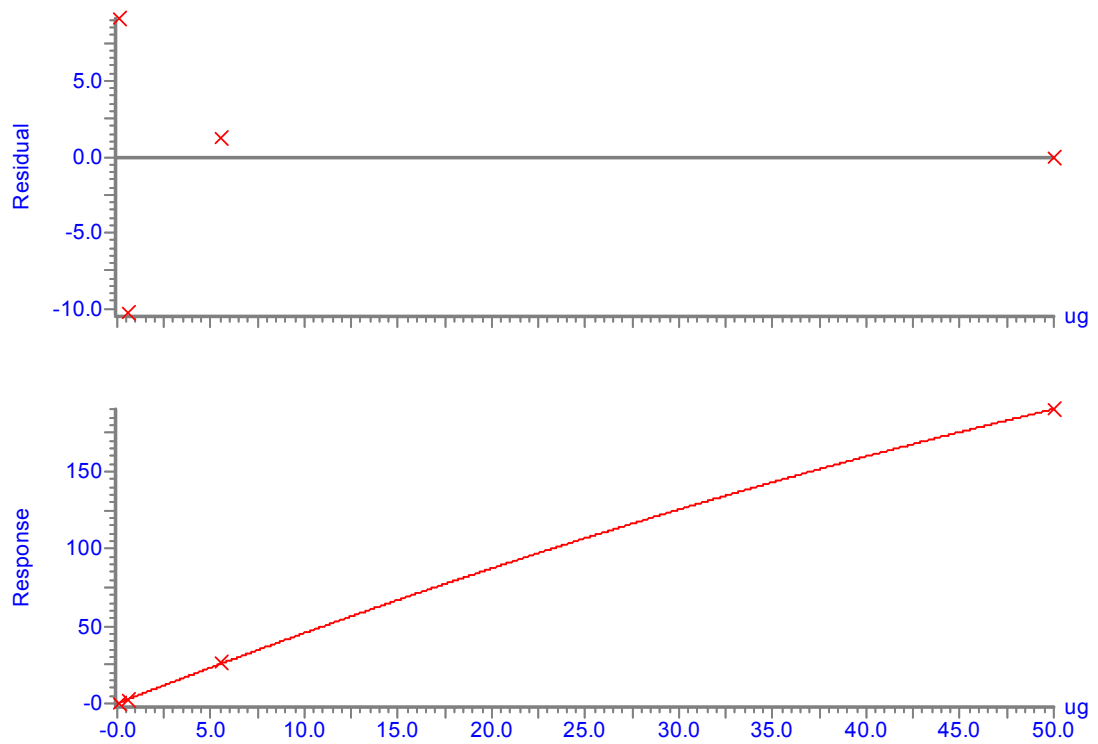
App. Figure 10. Calibration of LCMS using internal ^{13}C β-L-ODAP standards for the measurement of β-L-ODAP concentrations in batch 2 of the grass pea tissue samples

Batch 3

Calibration curve	$-0.0232021 \, x^2 + 5.26118 \, x - 0.00163096$
R^2	0.999905
Samples in this batch, unless otherwise indicated, all three replicate extractions were in this batch	LSWT11 – seedling shoot tip, seedling root tip 4946-7 – seedling shoot tip Pea – seedling shoot tip
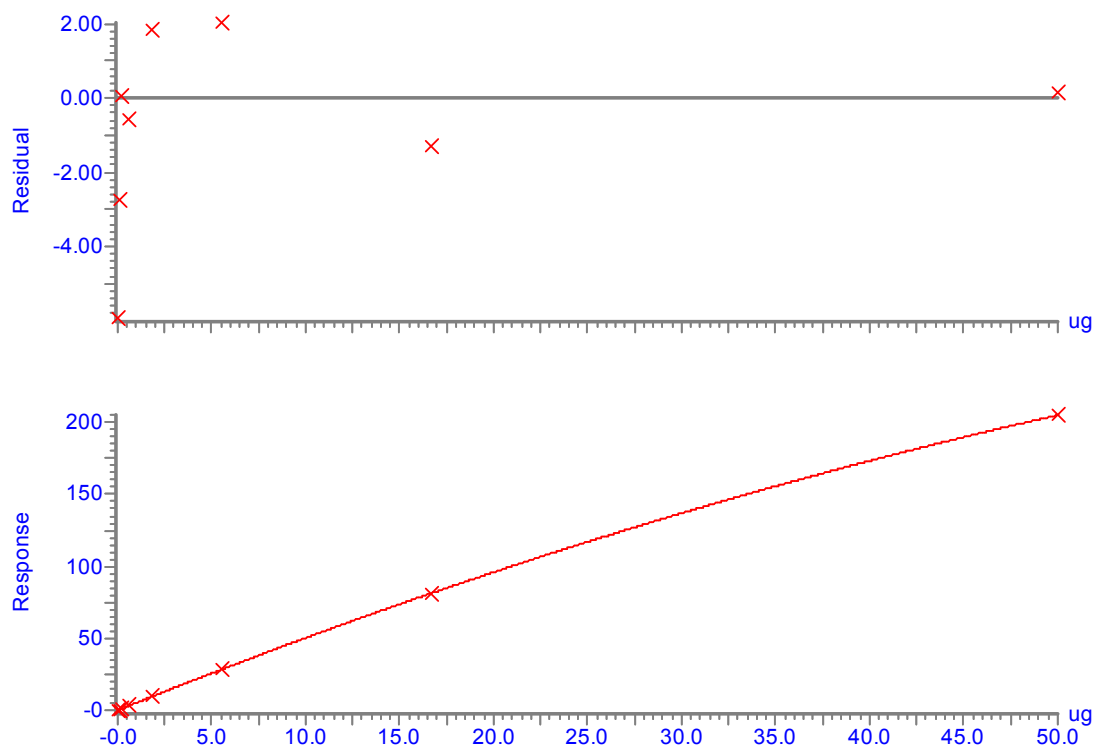
App. Figure 11. Calibration of LCMS using internal ^{13}C ß-L-ODAP standards for the measurement of ß-L-ODAP concentrations in batch 3 of the grass pea tissue samples

App. 3 Chapter 4 – Identification of candidate genes encoding metabolic enzymes in the β-L-ODAP biosynthetic pathway

App. 3.1 SortMeRNA command used for the processing of sequencing reads

```
sortmerna/1.9/src/sortmerna-1.9/scripts/merge-paired-reads.sh
trim_galore_out/*_R1_val_1.fq
trim_galore_out/*_R2_val_2.fq
merge_reads/LIB*_R1R2_trimmed_merged.fastq

sortmerna/1.9/src/sortmerna-1.9/sortmerna -I
merge_reads/LIB6*_R1R2_trimmed_merged.fastq -n 8 -db
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/rfam-5.8s-database-id98.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/rfam-5s-database-id98.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-arc-16s-database-id95.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-arc-23s-database-id98.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-bac-16s-database-id85.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-bac-23s-database-id98.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-euk-18s-database-id95.fasta
sortmerna/1.9/x86_64/sortmerna/rRNA_databases/silva-euk-28s-database-id98.fasta
        --accept
sortmerna/LIB*_R1R2.trimmed.merged.fastq.nonrRNA--other
sortmerna/LIB*_R1R2.trimmed.merged.fastq.rRNA --bydbs
        --log LIB6*.sortmerna.logfile --paired-out -r 0.25 -a 32 -v
```

App. 3.2 Trim Galore command

```
source   python-2.7.1;trim_galore/0.3.3/x86_64/bin/trim_galore   -q   20   --phred33   --
stringency 5 -length 60
        --paired --fastqc *_R1.fastq *_R2.fastq -o trim_galore_out
```

App. 3.3 Trinity assembly commands

```
trinityrnaseq-r2013_08_14/Trinity.pl  --seqType  fq  --JM  30G  --SS_lib_type  RF  --output
denovo_trinity_all_LIB_R1_trimmed_nonrna --CPU 16 --min_kmer_cov 2
        --bflyHeapSpaceMax 30G --bflyCPU 4 –left
trim_galore/sortmerna/unmerged_reads/all_LIB_R1_trimmed_nonrna.fastq--right
trim_galore/sortmerna/unmerged_reads/all_LIB_R2_trimmed_nonrna.fastq
```

App. 3.4 Commands used for open reading frame prediction and automatic annotation

```
transdecoder/r20131117/x86_64/bin/TransDecoder-t
        denovo_trinity_all_LIB_R1_trimmed_nonrna.fasta--MPI --CPU 10
annotF --fasta longest_orfss.pep
```

App. 3.5 Alignment of grass pea BAHD acyltransferases selected as ODAP-synthase candidates
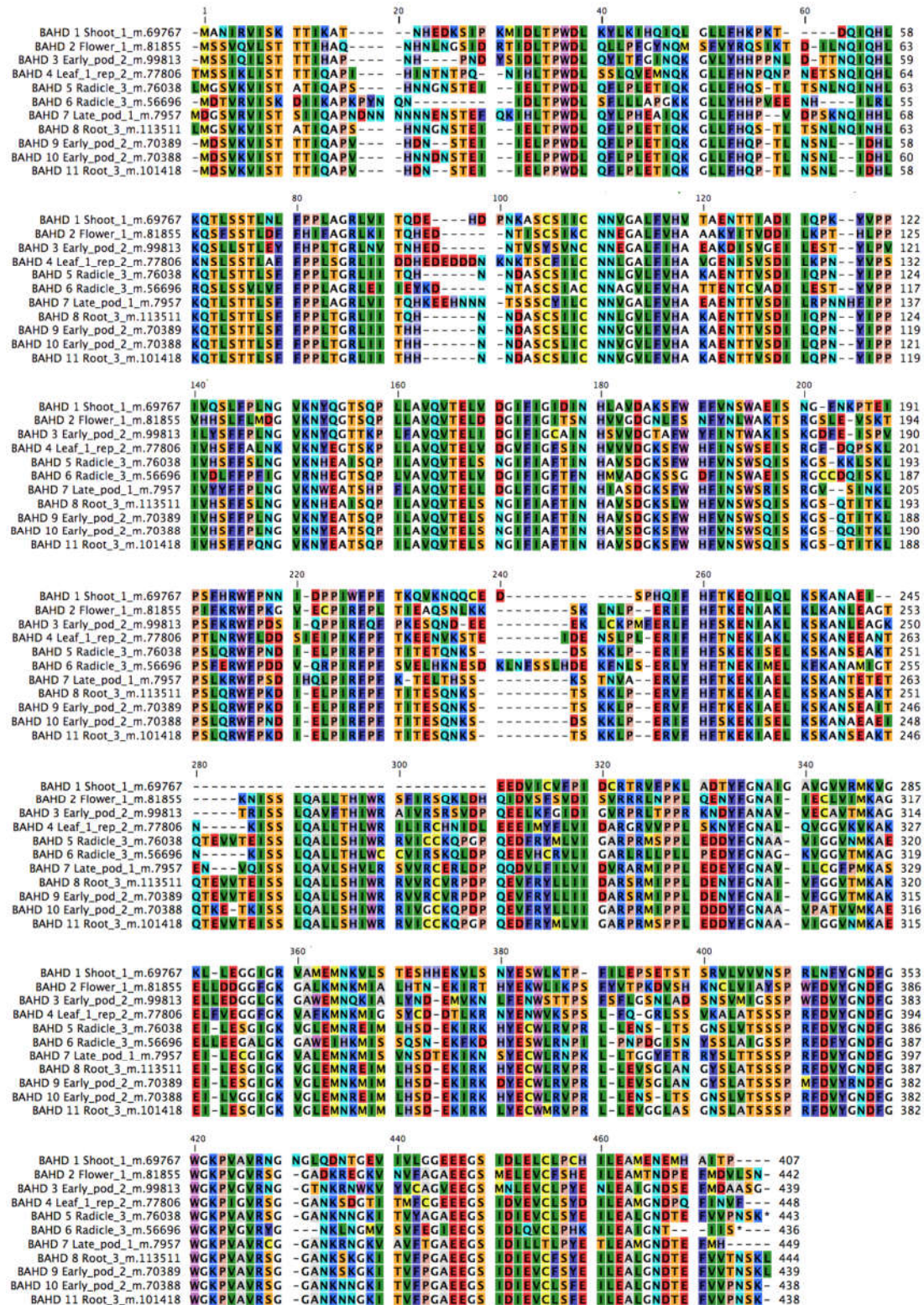
App. Figure 12. Alignment of predicted amino acid sequences of candidates for the ODAP-synthase, selected based on the grass pea transcriptomes. Alignment produced using CLC Main Workbench

App. 3.6 Phylogenetic tree of candidate clade of BAHD-ATs including related species
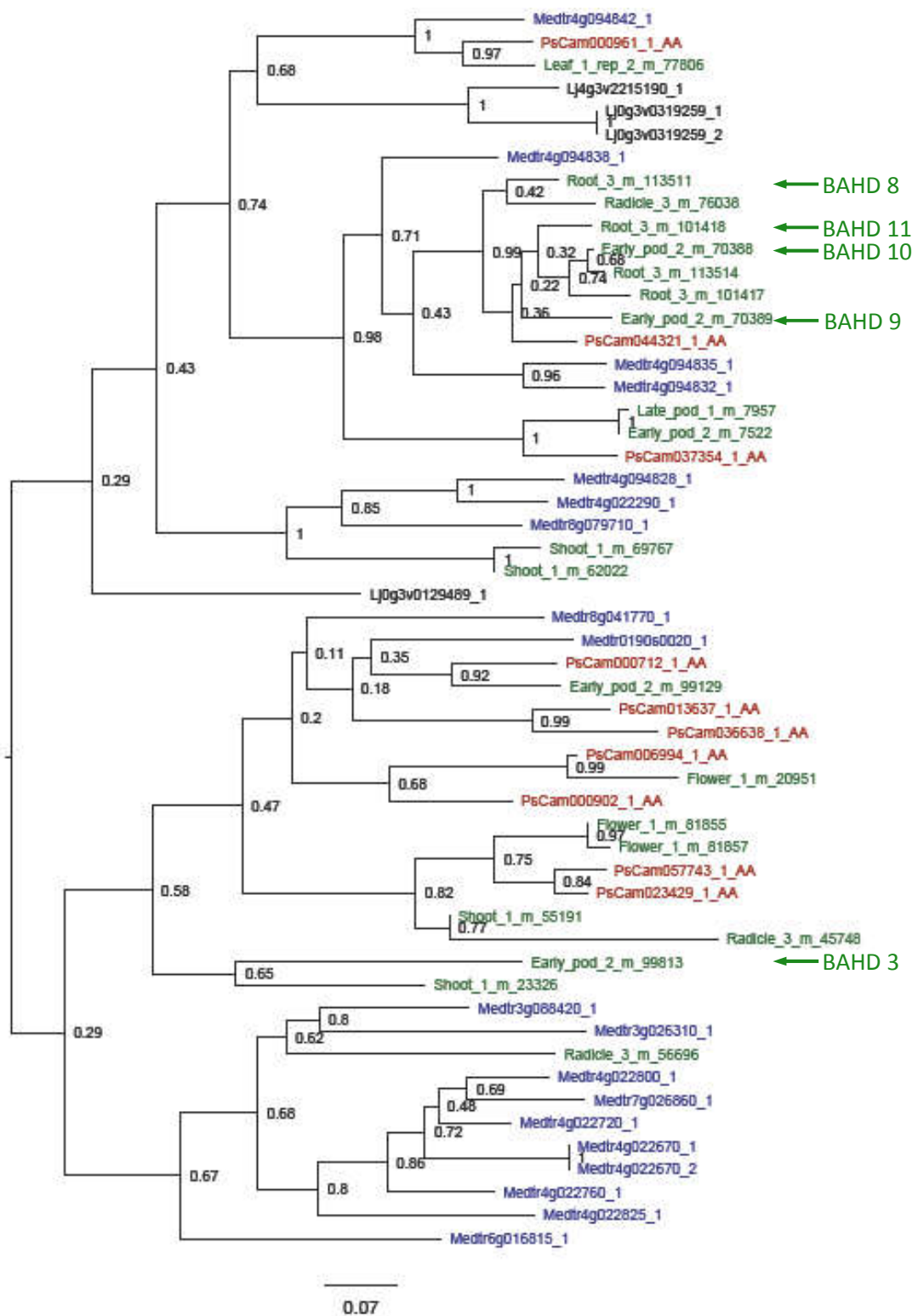
App. Figure 13. Phylogenetic tree showing the clade of BAHD-acyltransferases selected for further investigation. Nodes are coloured according to species: green - Lathyrus sativus; blue - Medicago truncatula; black - Lotus japonicus; red - Pisum sativum; Arabidopsis thaliana predicted proteins were included in the alignment, but none were grouped into this clade. Bootstrap values are shown at each branch. Grass pea proteins included in this phylogeny that are not shown in the phylogeny in section 4.3.5 were less than 400 amino acids long and had been excluded from the previous alignment. Candidate genes tested by heterologous expression are marked with arrows.

App. 3.7 Predicted change in protein structure caused by the deletion in the BAHD10 clone
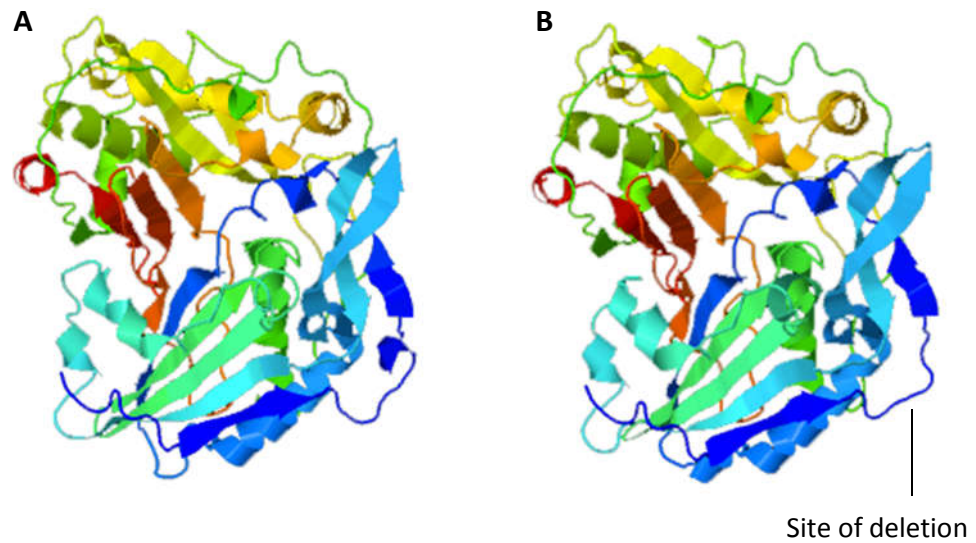
Site of deletion

App. Figure 14. Predicted tertiary structures of the predicted proteins of A) BAHD10 from the grass pea transcriptome and B) the cloned gene. The site of the two amino acid deletion is marked. The overall structure of the enzyme is not affected. Structure predictions calculated using Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/)

Bibliography

Abegaz, B.M., Nunn, P.B., De Bruyn, A., and Lambein, F. (1993). Thermal isomerization of *N*-oxalyl derivatives of diamino acids. Phytochemistry *33*, 1121-1123.

Addis, G., and Narayan, R. (2000). Interspecific hybridisation of *Lathyrus sativus* (Guaya) with wild *Lathyrus* species and embryo rescue. African Crop Science Journal *8*, 129-136.

Adsule, R.N., and Barat, G.K. (1977). Occurrence of oxalyl-CoA synthetase in Indian pulses. Experientia *33*, 416-417.

Akter, L., Mahbub, M., Habib, M.A., and Alam, S.S. (2015). Characterization of three varieties of *Lathyrus sativus* L. by fluorescent karyotype and RAPD analysis. Cytologia *80*, 457-465.

Aletor, V.A., El-Moneim, A.A., and Goodchild, A.V. (1994). Evaluation of the seeds of selected lines of three *Lathyrus* spp for β-N-oxalylamino-L-alanine (BOAA), tannins, trypsin inhibitor activity and certain in-vitro characteristics. Journal of the Science of Food and Agriculture *65*, 143-151.

Ali, L., Deokar, A., Caballo, C., Tar'an, B., Gil, J., Chen, W., Millan, T., and Rubio, J. (2016). Fine mapping for double podding gene in chickpea. Theoretical and Applied Genetics *129*, 77-86.

Almekinders, C.J., Louwaars, N.P., and De Bruijn, G. (1994). Local seed systems and their importance for an improved seed supply in developing countries. Euphytica *78*, 207-216.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.

Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A.-L., Jacquin, F., Klein, A., Martin, C., Boucherot, K., Kreplak, J.*, et al.* (2015). Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. The Plant Journal *84*, 1-19.

Anand, U. (2016). ICMR panel clears 'unsafe' khesari dal banned in 1961. In The Indian Express (New Delhi).

Arentoft, A.M.K., and Greirson, B.N. (1995). Analysis of 3-(N-oxalyl)-L-2,3-diaminopropanoic acid and its alpha-isomer in grass pea (*Lathyrus sativus*) by capillary zone electrophoresis. Journal of Agricultural and Food Chemistry *43*, 942-945.

Asmussen, C.B., and Liston, A. (1998). Chloroplast DNA Characters, Phylogeny, and Classification of Lathyrus (Fabaceae). American Journal of Botany *85*, 387-401.

Asthana, A. (1995). Grasspea cultivation in problem areas: present approaches. Paper presented at: Lathyrus Genetic Resources in Asia (Raipur, India: International Plant Genetic Resources Institute).

Bairu, M.W., Aremu, A.O., and Van Staden, J. (2011). Somaclonal variation in plants: causes and detection methods. Plant Growth Regulation *63*, 147-173.

Ballhorn, D.J., Kautz, S., Heil, M., and Hegeman, A.D. (2009). Analyzing plant defenses in nature. Plant signaling & behavior *4*, 743-745.

Bangladesh Bureau of Statistics (2016). Yearbook of Agricultural Statistics-2015. Ministry of Planning. (Dhaka:Statistics and Informatics Division). 27.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A.*, et al.* (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science *332*, 960-963.

Baranyk, P., and Fábry, A. (1999). History of the rapeseed (*Brassica napus* L.) growing and breeding from middle age Europe to Canberra. Paper presented at: International Rapeseed Congress.

Barik, D.P., Mohapatra, U., and Chand, P.K. (2005). Transgenic grasspea (*Lathyrus sativus* L.): factors influencing agrobacterium-mediated transformation and regeneration. Plant Cell Reports *24*, 523-531.

Barrow, M.V., Simpson, C.F., and Miller, E.J. (1974). Lathyrism: a review. Quarterly Review of Biology *49*, 101-128.

Bartley, L.E., Peck, M.L., Kim, S.R., Ebert, B., Manisseri, C., Chiniquy, D.M., Sykes, R., Gao, L., Rautengarten, C., Vega-Sanchez, M.E.*, et al.* (2013). Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. Plant Physiology *161*, 1615-1633.

Başaran, U., Hanife, M., Özlem, Ö.-A., Zeki, A., and İlknur, A. (2011). Variability in forage quality of Turkish grass pea (*Lathyrus sativus* L.) landraces. Turkish Journal of Field Crops *16*, 9-14.

Belaid, Y., Chtourou-Ghorbel, N., Marrakchi, M., and Trifi-Farah, N. (2006). Genetic diversity within and between populations of *Lathyrus* genus (Fabaceae) revealed by ISSR markers. Genetic Resources and Crop Evolution *53*, 1413-1418.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant methods *9*, 39.

Bell, E.A. (1964). Relevance of Biochemical Taxonomy to the Problem of Lathyrism. Nature *203*, 378-380.

Bell, E.A., Lackey, J.A., and Polhill, R.M. (1978). Systematic significance of canavanine in the Papilionoideae (Faboideae). Biochemical Systematics and Ecology *6*, 201-212.

Bell, E.A., Perera, C.K., Nunn, P.B., Simmonds, M.S., and Blaney, W.M. (1996). Non-protein amino acids of *Lathyrus latifolius* as feeding deterrents and phagostimulants in *Spodoptera littoralis*. Phytochemistry *43*, 1003-1007.

Berger, A., Meinhard, J., and Petersen, M. (2006). Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases. Planta *224*, 1503-1510.

Bett, C., Ouma, J.O., and Groote, H.D. (2010). Perspectives of gatekeepers in the Kenyan food industry towards genetically modified food. Food Policy *35*, 332-340.

Bhowmick, M.K. (2013). Productivity improvement of khesari (*Lathyrus sativus* L.) through low-cost technologies under rice-utera system in West Bengal. Proceedings of the 9th all India Peoples' technology congress.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764-764.

Bisignano, V., Gatta, C., and Polignano, G. (2002). Variation for protein content and seed weight in grass pea (*Lathyrus* spp.) germplasm. Plant Genetic Resources Newsletter, 30-34.

Bontpart, T., Cheynier, V., Ageorges, A., and Terrier, N. (2015). BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. New Phytologist *208*, 695-707.

Briggs, C., Parreno, N., and Campbell, C. (1983). Phytochemical assessment of *Lathyrus* species for the neurotoxic agent, beta-N-oxalyl-L-alpha-beta-diaminopropionic acid. Planta medica *47*, 188-190.

Butler, A., Tesfay, Z., D'Andrea, C., and Lyons, D. (1999). The ethnobotany of *Lathyrus sativus* L. in the highlands of Ethiopia. Book chapter in The Exploitation of Plant Resources in Ancient Africa, M. Van der Veen, ed. (Springer), pp. 123-136.

Byerlee, D., De Janvry, A., Sadoulet, E., Townsend, R., and Klytchnikova, I. (2007). World Development Report, 2008: agriculture for development. Washington: World Bank.

Campbell, C.G. (1997). Grass pea: *Lathyrus sativus* L, Vol 18 (International Plant Genetic Resources Institute).

Campbell, C.G., and Briggs, C.J. (1987). Registration of low neurotoxin content *Lathyrus* germplasm LS-8246. Crop Science *27*, 821-821.

Cantani, A. (1873). Latirismo (lathyrismus) illustrata da tre casi clinici. Morgagni, 745.

Cardoso, A.P., Mirione, E., Ernesto, M., Massaza, F., Cliff, J., Rezaul Haque, M., and Bradbury, J.H. (2005). Processing of cassava roots to remove cyanogens. Journal of Food Composition and Analysis *18*, 451-460.

Carpenter, J.E. (2010). Peer-reviewed surveys indicate positive impact of commercialized GM crops. Nature Biotechnology *28*, 319-321.

Cattivelli, L., Rizza, F., Badeck, F.-W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., Mare, C., Tondelli, A., and Stanca, A.M. (2008). Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crops Research *105*, 1-14.

Chakrabarti, A., Santha, I.M., and Mehta, S.L. (1999). Molecular characterisation of low ODAP somaclones of *Lathyrus sativus*. Journal of Plant Biochemistry and Biotechnology *8*, 25-29.

Chantreau, M., Grec, S., Gutierrez, L., Dalmais, M., Pineau, C., Demailly, H., Paysant-Leroux, C., Tavernier, R., Trouvé, J.-P., and Chatterjee, M. (2013). PT-Flax (phenotyping and TILLinG of flax): development of a flax (*Linum usitatissimum* L.) mutant population and TILLinG platform for forward and reverse genetics. BMC Plant Biology *13*, 1.

Chapman, M.A. (2015). Transcriptome sequencing and marker development for four underutilized legumes. Applications in Plant Sciences *3*, 1400111.

Chase, R.A., Pearson, S., Nunn, P.B., and Lantos, P.L. (1985). Comparative toxicities of α- and β-N-oxalyl-l-α,β-diaminopropionic acids to rat spinal cord. Neuroscience Letters *55*, 89-94.

Chiwona-Karltun, L., Mkumbira, J., Saka, J., Bovin, M., Mahungu, N.M., and Rosling, H. (1998). The importance of being bitter—a qualitative study on cassava cultivar preference in Malawi. Ecology of Food and Nutrition *37*, 219-245.

Chowdhury, M.A., and Slinkard, A.E. (2000). Genetic diversity in grasspea (*Lathyrus sativus* L.). Genetic Resources and Crop Evolution *47*, 163-169.

Chowdhury, M.A., and Sllnkard, A.E. (1997). Natural outcrossing in grasspea. Journal of Heredity *88*, 154-156.

Chowdhury, S., and Davis, R. (1989). Comparison of the effects of two lathyrogens on the reproductive system of the laying hen. The Veterinary Record *124*, 240-242.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Clancey, B. (2009). World Pulse Outlook. In Report to The Saskatchewan Pulse Growers (Stat Publishing), pp. 1-17.

Cohn, D., and Streifler, M. (1983). Intoxication by the chickling pea (*Lathyrus sativus*): nervous system and skeletal findings. Book chapter in Toxicology in the Use, Misuse, and Abuse of Food, Drugs, and Chemicals (Springer), pp. 190-193.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L., and Henikoff, S. (2001). High-throughput screening for induced point mutations. Plant Physiology *126*, 480-484.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674-3676.

Coward, F., Shennan, S., Colledge, S., Conolly, J., and Collard, M. (2008). The spread of Neolithic plant economies from the Near East to northwest Europe: a phylogenetic analysis. Journal of Archaeological Science *35*, 42-56.

D'Auria, J.C. (2006). Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology *9*, 331-340.

D'Auria, J.C., Chen, F., and Pichersky, E. (2002). Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. Plant Physiology *130*.

D'Auria, J.C., Pichersky, E., Schaub, A., Hansel, A., and Gershenzon, J. (2007a). Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)‐3‐hexen‐1‐yl acetate in *Arabidopsis thaliana*. The Plant Journal *49*, 194-207.

D'Auria, J.C., Reichelt, M., Luck, K., Svatoš, A., and Gershenzon, J. (2007b). Identification and characterization of the BAHD acyltransferase malonyl CoA:anthocyanidin 5-O-glucoside-6''-O-malonyltransferase (At5MAT) in *Arabidopsis thaliana*. FEBS Letters *581*.

Dai, A. (2013). Increasing drought under global warming in observations and models. Nature Climate Change *3*, 52-58.

Dasler, W., and Stoner, R. (1959). Production of spinal curvatures in rats by chemically dissimilar substances. Experientia *15*, 112-113.

Dastur, D.K., and Iyer, C.G.S. (1959). Lathyrism versus odoratism. Nutrition Reviews *17*, 33-36.

David, S., and Sperling, L. (1999). Improving technology delivery mechanisms: Lessons from bean seed systems research in eastern and central Africa. Agriculture and Human Values *16*, 381-388.

Dawson, D.A., Rinaldi, A.C., and Pöch, G. (2002). Biochemical and toxicological evaluation of agent-cofactor reactivity as a mechanism of action for osteolathyrism. Toxicology *177*, 267-284.

Dawson, D.A., Schuitz, T.W., Baker, L.L., and Mannar, A. (1990). Structure–activity relationships for osteolathyrism. III. Substituted thiosemicarbazides. Journal of Applied Toxicology *10*, 59-64.

De Bruyn, A., Becu, C., Lambein, F., Kebede, N., Abegaz, B., and B Nunn, P. (1994). The mechanism of the rearrangement of the neurotoxin β-ODAP to α-ODAP. Phytochemistry *36*, 85-89.

De Vega, J.J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J.L., Ergon, Å., Rognli, O.A., Jones, C., Swain, M., Geurts, R.*, et al.* (2015). Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. Scientific Reports *5*, 17394.

Deshpande, S.S., and Campbell, C.G. (1992). Genotype variation in BOAA, condensed tannins, phenolics and enzyme inhibitors of grass pea (*Lathyrus sativus*). Canadian journal of plant science *72*, 1037-1047.

Dixit, G.P., Parihar, A.K., Bohra, A., and Singh, N.P. (2016). Achievements and prospects of grass pea (*Lathyrus sativus* L.) improvement for sustainable food production. The Crop Journal *4*, 407-416.

Drouin, P., Prëvost, D., and Antoun, H. (2000). Physiological adaptation to low temperatures of strains of *Rhizobium leguminosarum* bv. *viciae* associated with *Lathyrus* spp. FEMS Microbiology Ecology *32*, 111-120.

Dudareva, N., D'Auria, J.C., Nam, K.H., Raguso, R.A., and Pichersky, E. (1998). Acetyl-CoA:benzylalcohol acetyltransferase - an enzyme involved in floral scent production in *Clarkia breweri*. Plant J *14*.

Dufour, D.L. (2011). Assessing diet in populations at risk for konzo and neurolathyrism. Food and Chemical Toxicology *49*, 655-661.

Duvenage, L., Hitzeroth, I.I., Meyers, A.E., and Rybicki, E.P. (2013). Expression in tobacco and purification of beak and feather disease virus capsid protein fused to elastin-like polypeptides. Journal of Virological Methods *191*, 55-62.

Dwivedi, M.P., and Prasad, B.G. (1964). An epidemiological study of lathyrism in the district of Rewa, Madhya Pradesh. Indian Journal of Medical Research *52*, 81-116.

Ebert, A. (2014). Potential of underutilized traditional vegetables and legume crops to contribute to food and nutritional security, income and more sustainable production systems. Sustainability *6*, 319.

El-Moneim, A.M.A., Van Dorrestein, B., Baum, M., Mulugeta, W., Center, P., and Debre Berhan, E. (1999). Role of ICARDA in improving the nutritional quality and yield potential of grasspea (*Lathyrus sativus* L.) for subsistence farmers in developing countries (International Food Policy Research Institute).

Engels, J., and Hawkes, J. (1991). The Ethiopian gene centre and its genetic diversity. Plant Genetic Resources of Ethiopia, Engels, J and J Hawkes (Eds) Cambridge University Press, New York, 23-26.

Enneking, D. (2011). The nutritive value of grasspea (*Lathyrus sativus*) and allied species, their toxicity to animals and the role of malnutrition in neurolathyrism. Food and Chemical Toxicology *49*, 694-709.

Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Pâques, F., and Lacroix, E. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Research *31*, 2952-2962.

Erskine, W., Smartt, J., and Muehlbauer, F.J. (1994). Mimicry of lentil and the domestication of common vetch and grass pea. Economic Botany *48*, 326-332.

Eslavath, R.K., Sharma, D., Bin Omar, N.A.M., Chikati, R., Teli, M.K., Rajanikant, G.K., and Singh, S.S. (2016). β-N-oxalyl-l-α, β- diaminopropionic acid induces HRE expression by inhibiting HIF-prolyl hydroxylase-2 in normoxic conditions. European Journal of Pharmacology *791*, 405-411.

Evans, C.S., and Bell, E.A. (1979). Non-protein amino acids of *Acacia* species and their effect on the feeding of the acridids *Anacridium melanorhodon* and *Locusta migratoria*. Phytochemistry *18*, 1807-1810.

FAO (2013). FAOStat world area harvested - primary oilcrops. UN Food and Agriculture Organisation. http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor. Accessed: 30/08/2016.

FAO/IAEA (2016). Joint FAO/IAEA Mutant Variety Database. FAO/IAEA. https://mvd.iaea.org/. Accessed: 18/01/2016.

Ferrio, J.P., Arab, G., Buxó, R., Guerrero, E., Molist, M., Voltas, J., and Araus, J.L. (2012). Agricultural expansion and settlement economy in Tell Halula (Mid-Euphrates valley): A diachronic study from early Neolithic to present. Journal of Arid Environments *86*, 104-112.

Fikre, A., Korbu, L., Kuo, Y.-H., and Lambein, F. (2008). The contents of the neuro-excitatory amino acid β-ODAP (β-N-oxalyl-l-α, β-diaminopropionic acid), and other free and protein amino acids in the seeds of different genotypes of grass pea (*Lathyrus sativus* L.). Food Chemistry *110*, 422-427.

Fikre, A., Lambein, F., and Gheysen, G. (2006). A life-saving food plant producing more neurotoxin under environmental stress. Commun Agric Appl Biol Sci *71*, 79-82.

Fikre, A., Negwo, T., Kuo, Y.H., Lambein, F., and Ahmed, S. (2011). Climatic, edaphic and altitudinal factors affecting yield and toxicity of *Lathyrus sativus* grown at five locations in Ethiopia. Food and Chemical Toxicology *49*, 623-630.

Foster, J., Kim, H.U., Nakata, P.A., and Browse, J. (2012). A previously unknown oxalyl-CoA synthetase is important for oxalate catabolism in *Arabidopsis*. The Plant Cell Online *24*, 1217-1229.

Foster, J., Luo, B., and Nakata, P.A. (2016). An oxalyl-CoA dependent pathway of oxalate catabolism plays a role in regulating calcium oxalate crystal accumulation and defending against oxalate-secreting phytopathogens in *Medicago truncatula*. Plos One *11*, e0149850.

Foster, J., and Nakata, P.A. (2014). An oxalyl-CoA synthetase is important for oxalate metabolism in *Saccharomyces cerevisiae*. FEBS Letters *588*, 160-166.

Fox, J.L. (2013). Mars collaborates to sequence Africa's neglected food crops. In Nature Biotechnolgy (Nature Publishing Group), pp. 867-867.

Foyer, C.H., Lam, H.-M., Nguyen, H.T., Siddique, K.H.M., Varshney, R.K., Colmer, T.D., Cowling, W., Bramley, H., Mori, T.A., Hodgson, J.M.*, et al.* (2016). Neglecting legumes has compromised human health and sustainable food production. Nature Plants *2*, 16112.

Franceschi, V.R., and Nakata, P.A. (2005). Calcium oxalate in plants: formation and function. Annual Review of Plant Biology *56*, 41-71.

FSSAI (2011). Food Safety and Standards (Prohibition and restriction on sales) regulations. Food Safety Standards Authority of India. (New Delhi:The Gazette of India). 3:10-16.

Gannon, R.L., and Terrian, D.M. (1989). BOAA selectively enhances L-glutamate release from guinea pig hippocampal mossy fiber synaptosomes. Neuroscience Letters *107*, 289-294.

Garcia, J., Barker, D.G., and Journet, E.-P. (2006). Seed storage and germination. Book chapter in The Medicago truncatula Handbook (Ardmore, Oklahoma: The Samuel Roberts Noble Foundation).

Gassert, F. (2013). One-Quarter of World's Agriculture Grows in Highly Water-Stressed Areas. World Resources Institute. http://www.wri.org/blog/2013/10/one-quarter-world%E2%80%99s-agriculture-grows-highly-water-stressed-areas. Accessed: 02/09/2016.

Getahun, H., Mekonnen, A., TekleHaimanot, R., and Lambein, F. (1999). Epidemic of neurolathyrism in Ethiopia. The Lancet *354*, 306-307.

Ghasem, K., Danesh-Gilevaei, M., and Aghaalikhani, M. (2011). Karyotypic and nuclear DNA variations in *Lathyrus sativus* (Fabaceae). Caryologia *64*, 42-54.

Ghosh, B., Mitra, J., Chakraborty, S., Bhattacharyya, J., Chakraborty, A., Sen, S.K., and Neerathilingam, M. (2015). Simple detection methods for antinutritive factor β-ODAP present in *Lathyrus sativus* L. by high pressure liquid chromatography and thin layer chromatography. Plos One *10*, e0140649.

Girma, A., Tefera, B., and Dadi, L. (2011). Grass pea and neurolathyrism: farmers' perception on its consumption and protective measure in North Shewa, Ethiopia. Food and Chemical Toxicology *49*, 668-672.

Girma, D., and Korbu, L. (2012). Genetic improvement of grass pea (*Lathyrus sativus*) in Ethiopia: an unfulfilled promise. Plant Breeding *131*, 231-236.

Godwin, I., Todd, G., Ford-Lloyd, B., and Newbury, H.J. (1991). The effects of acetosyringone and pH on *Agrobacterium*-mediated transformation vary according to plant species. Plant Cell Reports *9*, 671-675.

GPC (2016). United Nations Proclaims 2016 as "International Year of Pulses". Global Pulse Confederation. http://iyp2016.org/about-us. Accessed: 01/09/2016.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q*., et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology *29*, 644-652.

Greilhuber, J., and Ebert, I. (1994). Genome size variation in *Pisum sativum*. Genome *37*, 646-655.

Gutiérrez-Marcos, J.F., Vaquero, F., Sáenz de Miera, L.E., and Vences, F.J. (2006). High genetic diversity in a world-wide collection of *Lathyrus sativus* L. revealed by isozymatic analysis. Plant Genetic Resources *4*, 159-171.

Haimanot, R., Kidane, Y., Wuhib, E., Kassina, A., Endeshaw, Y., Alemu, T., and Spencer, P. (1993). The epidemiology of lathyrism in north and central Ethiopia. Ethiopian Medical Journal *31*, 15-24.

Haimanot, R.T., Feleke, A., and Lambein, F. (2005). Is lathyrism still endemic in northern Ethiopia?–The case of Legambo Woreda (district) in the South Wollo Zone, Amhara National Regional State. Ethiopian Journal of Health Development *19*, 230-236.

Haimanot, R.T., Kidane, Y., Wuhib, E., Kalissa, A., Alemu, T., Zein, Z.A., and Spencer, P.S. (1990). Lathyrism in rural northwestern ethiopia: a highly prevalent neurotoxic disorder. International Journal of Epidemiology *19*, 664-672.

Hanbury, C., White, C., Mullan, B., and Siddique, K. (2000). A review of the potential of *Lathyrus sativus* L. and *L. cicera* L. grain for use as animal feed. Animal Feed Science and Technology *87*, 1-27.

Haque, A., Hossain, M., Lambein, F., and Bell, E.A. (1997). Evidence of osteolathyrism among patients suffering from neurolathyrism in Bangladesh. Natural Toxins *5*, 43-46.

Haque, A., Hossain, M., Wouters, G., and Lambein, F. (1996). Epidemiological study of lathyrism in northwestern districts of Bangladesh. Neuroepidemiology *15*, 83-91.

Haque, R.M., Kuo, Y.H., Lambein, F., and Hussain, M. (2011). Effect of environmental factors on the biosynthesis of the neuro-excitatory amino acid beta-ODAP (beta-N-oxalyl-L-alpha,beta-diaminopropionic acid) in callus tissue of *Lathyrus sativus*. Food and Chemical Toxicology *49*, 583-588.

Harrison, F.L., Nunn, P.B., and Hill, R.R. (1977). Synthesis of α- and β-N-oxalyl-l-α,β-diaminopropionic acids and their isolation from seeds of *Lathyrus sativus*. Phytochemistry *16*, 1211-1215.

Hendley, A.G. (1903). Lathyrism. The British Medical Journal *2*, 707-709.

Henikoff, S., Till, B.J., and Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. Plant Physiology *135*, 630-636.

Heringa, R.J., van Norel, A., and Tazelaar, M.F. (1969). Resistance to powdery mildew (*Erisyphe polygoni* D.C.) in peas (*Pisum sativum* L.). Euphytica *18*, 163-169.

Hillocks, R., and Maruthi, M. (2012). Grass pea (Lathyrus sativus): Is there a case for further crop improvement? Euphytica *186*, 647-654.

Holbrook, T.C., Gilliam, L.L., Stein, F.P., Morgan, S.E., Avery, A.L., Confer, A.W., and Panciera, R.J. (2015). Lathyrus hirsutus (*Caley Pea*) intoxication in a herd of horses. Journal of Veterinary Internal Medicine *29*, 294-298.

Hole, C., and Hardwick, R. (1976). Development and control of the number of flowers per node in *Pisum sativum* L. Annals of Botany *40*, 707-722.

Hu, H., and Xiong, L. (2014). Genetic engineering and breeding of drought-resistant crops. Annual Review of Plant Biology *65*, 715-741.

Huang, L.S., and Sternberg, P.W. (2006). Genetic dissection of developmental pathways. Book chapter in WormBook (The *C. elegans* Research Community).

Hussain, M. (1989). Future of khesari cultivation in Bangladesh. Advances in Pulses Research in Bangladesh, 183.

Hussain, M., Chowdhurry, B., Haque, R., Wouters, G., and Campbell, C. (1994). A comparative study of the O‐phthalaldehyde method for the neurotoxin 3‐N‐oxalyl‐l‐2, 3‐diaminopropanoic acid as modified by various laboratories. Phytochemical Analysis *5*, 247-250.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature Biotechnology *31*, 227-229.

Ikegami, F., Itagaki, S., Ishikawa, T., Ongena, G., Kuo, Y.H., Lambein, F., and Murakoshi, I. (1991). Biosynthesis of beta-(isoxazolin-5-on-2-yl)alanine, the precursor of the neurotoxic amino acid beta-N-oxalyl-L-alpha,beta-diaminopropionic acid. Chemical & Pharmaceutical Bulletin *39*, 3376-3377.

Ikegami, F., Lambein, F., Kuo, Y.-H., and Murakoshi, I. (1984). Isoxazolin-5-one derivatives in *Lathyrus odoratus* during development and growth. Phytochemistry *23*, 1567-1569.

Ikegami, F., Ongena, G., Sakai, R., Itagaki, S., Kobori, M., Ishikawa, T., Kuo, Y.-H., Lambein, F., and Murakoshi, I. (1993). Biosynthesis of β-(isoxazolin-5-on-2-yl)-l-alanine by cysteine synthase in *Lathyrus sativus*. Phytochemistry *33*, 93-98.

Ikegami, F., Yamamoto, A., Kuo, Y.H., and Lambein, F. (1999). Enzymatic formation of 2,3-diaminopropionic acid, the direct precursor of the neurotoxin beta-ODAP, in *Lathyrus sativus*. Biological & Pharmaceutical Bulletin *22*, 770-771.

Iqbal, A., Khalil, I.A., Ateeq, N., and Sayyar Khan, M. (2006). Nutritional quality of important food legumes. Food Chemistry *97*, 331-335.

Jackson, M.T., and Yunus, A.G. (1984). Variation in the grass pea (*Lathyrus sativus* L.) and wild species. Euphytica *33*, 549-559.

Jaoui, M., Kleindienst, T.E., Lewandowski, M., and Edney, E.O. (2004). Identification and quantification of aerosol polar oxygenated compounds bearing carboxylic or hydroxyl groups. Anal Chem *76*, 4765-4778.

Jiang, J., Su, M., Chen, Y., Gao, N., Jiao, C., Sun, Z., Li, F., and Wang, C. (2013). Correlation of drought resistance in grass pea (*Lathyrus sativus*) with reactive oxygen species scavenging and osmotic adjustment. Biologia *68*, 231-240.

Jiao, C.-J., Xu, Q.-L., Wang, C.-Y., Li, F.-M., Li, Z.-X., and Wang, Y.-F. (2006). Accumulation pattern of toxin β-ODAP during lifespan and effect of nutrient elements on β-ODAP content in *Lathyrus sativus* seedlings. The Journal of Agricultural Science *144*, 369-375.

Jiao, C.J., Jiang, J.L., Ke, L.M., Cheng, W., Li, F.M., Li, Z.X., and Wang, C.Y. (2011a). Factors affecting beta-ODAP content in *Lathyrus sativus* and their possible physiological mechanisms. Food and Chemical Toxicology *49*, 543-549.

Jiao, C.J., Jiang, J.L., Li, C., Ke, L.M., Cheng, W., Li, F.M., Li, Z.X., and Wang, C.Y. (2011b). beta-ODAP accumulation could be related to low levels of superoxide anion and hydrogen peroxide in *Lathyrus sativus* L. Food and Chemical Toxicology *49*, 556-562.

Jones, H.D. (2015) Challenging regulations: Managing risks in crop biotechnology. Food and Energy Security *4*, 87-91

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G.*, et al.* (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236-1240.

Jönsson, R. (1977). Erucic‐acid heredity in rapeseed (*Brassica napus* L. and *Brassica campestris* L.). Hereditas *86*, 159-170.

Joshi, P.K., and Rao, P.P. (2016). Global and regional pulse economies: Current trends and outlook, Vol 144 (International Food Policy Research Institute).

Jyothi, P., Rudra, M.P., and Rao, S.L. (1998). In vivo metabolism of beta-N-oxalyl-L-alpha,beta-diaminopropionic acid: the Lathyrus sativus neurotoxin in experimental animals. Natural Toxins *6*, 189-195.

Kanagarajan, S., Muthusamy, S., Gliszczyńska, A., Lundgren, A., and Brodelius, P.E. (2012). Functional expression and characterization of sesquiterpene synthases from *Artemisia annua* L. using transient expression system in *Nicotiana benthamiana*. Plant Cell Reports *31*, 1309-1319.

Kanchiswamy, C.N., Malnoy, M., Velasco, R., Kim, J.-S., and Viola, R. (2015). Non-GMO genetically edited crop plants. Trends in Biotechnology *33*, 489-491.

Kapila, J., De Rycke, R., Van Montagu, M., and Angenon, G. (1997). An *Agrobacterium*-mediated transient gene expression system for intact leaves. Plant Science *122*, 101-108.

Karadag, Y., and Yavuz, M. (2010). Seed yields and biochemical compounds of grasspea (*Lathyrus sativus* L.) lines grown in semi-arid regions of Turkey. African Journal of Biotechnology *9*, 8343.

Khan, J.K., Kuo, Y.H., Haque, A., and Lambein, F. (1995). Inhibitory and excitatory amino acids in cerebrospinal fluid of neurolathyrism patients, a highly prevalent motorneurone disease. Acta Neurol Scand *91*, 506-510.

Kharkwal, M. (1998). Induced mutations in chickpea (*Cicer arietinum* L.) I. comparative mutagenic effectiveness and efficiency of physical & chemical mutagens. The Indian Journal of Genetics and Plant Breeding *58*, 159-167.

Khoury, C.K., Bjorkman, A.D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L.H., and Struik, P.C. (2014). Increasing homogeneity in global food supplies and the implications for food security. Proceedings of the National Academy of Sciences *111*, 4001-4006.

Kim, K.S., Min, J.-Y., and Dickman, M.B. (2008). Oxalic acid is an elicitor of plant programmed cell death during *Sclerotinia sclerotiorum* disease development. Molecular Plant-Microbe Interactions *21*, 605-612.

Kislev, M.E. (1989). Origins of the cultivation of *Lathyrus sativus* and *L. cicera* (Fabaceae). Economic Botany *43*, 262-270.

Koh, H.L., Lau, A.J., and Chan, E.C.Y. (2005). Hydrophilic interaction liquid chromatography with tandem mass spectrometry for the determination of underivatized dencichine (β-N-oxalyl-L-α, β-diaminopropionic acid) in *Panax* medicinal plant species. Rapid communications in mass spectrometry *19*, 1237-1244.

Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Molecular and General Genetics MGG *204*, 383-396.

Kondra, Z., and Stefansson, B. (1970). Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. Canadian journal of plant science *50*, 643-647.

Kosma, D.K., Molina, I., Ohlrogge, J.B., and Pollard, M. (2012). Identification of an *Arabidopsis* Fatty Alcohol:Caffeoyl-Coenzyme A Acyltransferase Required for the Synthesis of Alkyl Hydroxycinnamates in Root Waxes. Plant Physiology *160*, 237-248.

Krausmann, F., Gingrich, S., Eisenmenger, N., Erb, K.-H., Haberl, H., and Fischer-Kowalski, M. (2009). Growth in global materials use, GDP and population during the 20th century. Ecological Economics *68*, 2696-2705.

Krishna, V.V., and Qaim, M. (2008). Potential impacts of Bt eggplant on economic surplus and farmers' health in India. Agricultural Economics *38*, 167-180.

Kumar, A.P.K., McKeown, P.C., Boualem, A., Ryder, P., Brychkova, G., Bendahmane, A., Sarkar, A., Chatterjee, M., Spillane, C. (2017) Molecular Breeding *37*, 14

Kumar, S., Bejiga, G., Ahmed, S., Nakkoul, H., and Sarker, A. (2011a). Genetic improvement of grass pea for low neurotoxin (beta-ODAP) content. Food and Chemical Toxicology *49*, 589-600.

Kumar, S., Imtiaz, M., Gupta, S., and Pratap, A. (2011b). Distant hybridization and alien gene introgression. Biology and breeding of food legumes, 81-110.

Kumar, V., Chattopadhyay, A., Ghosh, S., Irfan, M., Chakraborty, N., Chakraborty, S., and Datta, A. (2016). Improving nutritional quality and fungal tolerance in soya bean and grass pea by expressing an oxalate decarboxylase. Plant Biotechnology Journal *14*, 1394-1405.

Kuo, Y.-H., Bau, H.-M., Quemener, B., Khan, J.K., and Lambein, F. (1995). Solid-state fermentation of Lathyrus sativus seeds using Aspergillus oryzae and Rhizopus oligosporus sp T-3 to eliminate the neurotoxin β-ODAP without loss of nutritional value. Journal of the Science of Food and Agriculture *69*, 81-89.

Kuo, Y.-H., Ikegami, F., and Lambein, F. (1998). Metabolic routes of β-(isoxazolin-5-on-2-yl)-L-alanine (BIA), the precursor of the neurotoxin ODAP (β-N-oxalyl-L-α, β,-diaminopropionic acid), in different legume seedlings. Phytochemistry *49*, 43-48.

Kuo, Y.-H., Ikegami, F., and Lambein, F. (2003). Neuroactive and other free amino acids in seed and young plants of *Panax ginseng*. Phytochemistry *62*, 1087-1091.

Kuo, Y.-H., Khan, J.K., and Lambein, F. (1994). Biosynthesis of the neurotoxin β-ODAP in developing pods of *Lathyrus sativus*. Phytochemistry *35*, 911-913.

Kuo, Y.-H., and Lambein, F. (1991). Biosynthesis of the neurotoxin β-N-oxalyl-α, β-diaminopropionic acid in callus tissue of *Lathyrus sativus*. Phytochemistry *30*, 3241-3244.

Kusama-Eguchi, K., Ikegami, F., Kusama, T., Suda, A., Ogawa, Y., Igarashi, K., and Watanabe, K. (2005). A rat model of neurolathyrism: repeated injection of L-beta-ODAP induces the paraparesis of the hind legs. Amino Acids *28*, 139-143.

Kusama-Eguchi, K., Yamazaki, Y., Ueda, T., Suda, A., Hirayama, Y., Ikegami, F., Watanabe, K., May, M., Lambein, F., and Kusama, T. (2010). Hind-limb paraparesis in a rat model for neurolathyrism associated with apoptosis and an impaired vascular endothelial growth factor system in the spinal cord. Journal of Comparative Neurology *518*, 928-942.

Kusama-Eguchi, K., Yoshino, N., Minoura, A., Watanabe, K., Kusama, T., Lambein, F., and Ikegami, F. (2011). Sulfur amino acids deficiency caused by grass pea diet plays an important role in the toxicity of L-beta-ODAP by increasing the oxidative stress: studies on a motor neuron cell line. Food and Chemical Toxicology *49*, 636-643.

Lakshmanan, J., Cheema, P., and Padmanaban, G. (1971). Mechanism of action of β-N-oxalyl-L-α, β-diaminopropionic acid, *Lathyrus sativus* neurotoxin: Effect on brain lysosomes. Nature *234*, 156-157.

Lambein, F. (2000). Homeopathy, longevity and *Lathyrus sativus* toxicity. Lathyrus lathyrism Newsletter *1*, 4-5.

Lambein, F., Godelieve, O., and Yu-Haey, K. (1990). β-Isoxazolinone-alanine is involved in the biosynthesis of the neurotoxin β-N-oxalyl-l-α, β-diaminopropionic acid. Phytochemistry *29*, 3793-3796.

Lambein, F., Haque, R., Khan, J.K., Kebede, N., and Kuo, Y.H. (1994). From soil to brain: zinc deficiency increases the neurotoxicity of *Lathyrus sativus* and may affect the susceptibility for the motorneurone disease neurolathyrism. Toxicon *32*, 461-466.

Lambein, F., and Kuo, Y.-H. (2013). *Lathyrus sativus*: to eat or not to eat? CCDN NEWS, 1-2.

Lambein, F., Ngudi, D.D., and Kuo, Y.-H. (2001). Vapniarca revisited: Lessons from an inhuman human experience. Lathyrus lathyrism Newsletter, 5-7.

Lan, G., Lan, F., and Sun, X. (2016). Use of dencichine in preparation of drug for treating thrombocytopenia (Google Patents).

Lane, B., Dunwell, J.M., Ray, J., Schmitt, M., and Cuming, A. (1993). Germin, a protein marker of early plant development, is an oxalate oxidase. Journal of Biological Chemistry *268*, 12239-12242.

Langridge, P., and Reynolds, M.P. (2015). Genomic tools to assist breeding for drought tolerance. Current Opinion in Biotechnology *32*, 130-135.

Ledford, H. (2016). Gene-editing surges as US rethinks regulations. In Nature News (Nature).

Lees, S., Eyre, D.R., and Barnard, S.M. (1990). BAPN dose dependence of mature crosslinking in bone matrix collagen of rabbit compact bone: corresponding variation of sonic velocity and equatorial diffraction spacing. Connective Tissue Research *24*, 95-105.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.

Lioi, L., and Galasso, I. (2013). Development of genomic simple sequence repeat markers from an enriched genomic library of grass pea (*Lathyrus sativus* L.). Plant Breeding *132*, 649-653.

Livingstone, D.M., Hampton, J.L., Phipps, P.M., and Grabau, E.A. (2005). Enhancing resistance to *Sclerotinia minor* in peanut by expressing a barley oxalate oxidase gene. Plant Physiology *137*, 1354-1362.

Lohr, K., Camacho, A., and Vernooy, R. (2015). Seed systems–an overview. Book chapter in Farmers' seed systems: the challenge of linking formal and informal seed systems (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH), pp. 3-7.

Louwaars, N.P., and De Boef, W.S. (2012). Integrated seed sector development in Africa: a conceptual framework for creating coherence between practices, programs, and policies. Journal of Crop Improvement *26*, 39-59.

Louwaars, N.P., de Boef, W.S., and Edeme, J. (2013). Integrated seed sector development in Africa: a basis for seed policy and law. Journal of Crop Improvement *27*, 186-214.

Luo, J., Fuell, C., Parr, A., Hill, L., Bailey, P., Elliott, K., Fairhurst, S.A., Martin, C., and Michael, A.J. (2009). A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed. The Plant Cell *21*, 318-333.

Ma, X., Koepke, J., Panjikar, S., Fritzsch, G., and Stockigt, J. (2005). Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. Journal of Biological Chemistry *280*, 13576-13583.

Mahler-Slasky, Y., and Kislev, M.E. (2010). *Lathyrus* consumption in late Bronze and iron age sites in Israel: an Aegean affinity. Journal of Archaeological Science *37*, 2477-2485.

Malathi, K., Padmanab.G, and Sarma, P.S. (1970). Biosynthesis of beta-N-oxalyl-L-alpha,beta-diaminopropionic acid, *Lathyrus sativus* neurotoxin. Phytochemistry *9*, 1603-1609.

Malathi, K., Padmanaban, G., Rao, S.L., and Sarma, P.S. (1967). Studies on the biosynthesis of beta-N-oxalyl-L-alpha, beta-diaminopropionic acid, the *Lathyrus sativus* neurotoxin. Biochim Biophys Acta *141*, 71-78.

Malathi, K., Padmanaban, G., and Sarma, P.S. (1968). Oxalylation of some amino acids by an enzyme preparation from *Lathyrus sativus*. Indian J Biochem *5*, 184-185.

Malek, M., and Gazipur, B. (1999). Genetic resources of grass pea (*Lathyrus sativus* L.) in Bangladesh. Lathyrus Genetic Resources Network *8*, 1.

Manjaya, J.G. (2009). Genetic improvement of soybean variety VLS-2 through induced mutations. Paper presented at: Induced Plant Mutations in the Genomics Era (Vienna, Austria: Joint FAO/IAEA Programme).

Maple, J., and Møller, S.G. (2007). Mutagenesis in arabidopsis. Circadian Rhythms: Methods and Protocols, 197-206.

Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V., and Gleba, Y. (2005). Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants. Nature Biotechnology *23*, 718-723.

Marinova, E. (2007). Archaeobotanical data from the early Neolithic of Bulgaria. Book chapter in The origins and spread of domestic plants in southwest Asia and Europe, S. Colledge, and J. Conolly, eds. (Institute of Archeology, University College, London), pp. 93-109.

Massawe, F., Mayes, S., and Cheng, A. (2016). Crop Diversity: An Unexploited Treasure Trove for Food Security. Trends in Plant Science *21*, 365-368.

Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M.*, et al.* (2014). Data access for the 1,000 Plants (1KP) project. GigaScience *3*, 1-10.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nature Biotechnology *18*, 455-457.

McHughen, A., and Smyth, S. (2008). US regulatory system for genetically modified [genetically modified organism (GMO), rDNA or transgenic] crop cultivars. Plant Biotechnology Journal *6*, 2-12.

Mehta, T., Parker, A.J., Cusick, P.K., Zarghami, N.S., and Haskell, B.E. (1980). The *Lathyrus sativus* neurotoxin: evidence of selective retention in monkey tissue. Toxicology and Applied Pharmacology *52*, 54-61.

Mehta, T., Parker, A.J., Cusick, P.K., Zarghami, N.S., and Haskell, B.E. (1983). The *Lathyrus sativus* neurotoxin: resistance of the squirrel monkey to prolonged oral high doses. Toxicology and Applied Pharmacology *69*, 480-484.

Mehta, T., Zarghami, N.S., Cusick, P.K., Parker, A.J., and Haskell, B.E. (1976). Tissue distribution and metabolism of the *Lathyrus sativus* neurotoxin, L-3-oxalylamino-2-aminopropionic acid, in the squirrel monkey. Journal of Neurochemistry *27*, 1327-1331.

Meiner, Z., and Gotkine, M. (2016). Using advanced imaging methods to study neurolathyrism. The Israel Medical Association Journal *18*, 341-345.

Membre, N., Berna, A., Neutelings, G., David, A., David, H., Staiger, D., Saez Vasquez, J., Raynal, M., Delseny, M., and Bernier, F. (1997). cDNA sequence, genomic organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. Plant Molecular Biology *35*, 459-469.

Meyer, R.S., DuVal, A.E., and Jensen, H.R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytologist *196*, 29-48.

Miah, M.M., and Haque, A. (2013). Policy options for supporting agricultural diversification in Bangladesh. Final report on the NFPCSP/FAO commissioned study Food Planning and Monitoring Unit, Ministry of Food, Dhaka, Bangladesh.

Mikic, A., Mihailovic, V., Cupina, B., Duric, B., Krstic, D., Vasic, M., Vasiljevic, S., Karagic, D., and Dordevic, V. (2011). Towards the re-introduction of grass pea (*Lathyrus sativus*) in the West Balkan countries: the case of Serbia and Srpska (Bosnia and Herzegovina). Food and Chemical Toxicology *49*, 650-654.

Milczak, M., Pedzinski, M., Mnichowska, H., Szwed-Urbas, K., and Rybinski, W. (2001). Creative breeding of grasspea (*Lathyrus sativus* L.) in Poland. Lathyrus lathyrism Newsletter *2*, 85-88.

Milkowski, C., and Strack, D. (2004). Serine carboxypeptidase-like acyltransferases. Phytochemistry *65*, 517-524.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.-L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A.*, et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology *25*, 778-785.

Murti, V., Seshadri, T., and Venkitasubramanian, T. (1964). Neurotoxic compounds of the seeds of *Lathyrus sativus*. Phytochemistry *3*, 73-78.

Nandini, A.V., Murray, B.G., O'Brien, I.E.W., and Hammett, K.R.W. (1997). Intra- and interspecific variation in genome size in *Lathyrus* (Leguminosae). Botanical Journal of the Linnean Society *125*, 359-366.

Narayan, R. (1998). The role of genomic constraints upon evolutionary changes in genome size and chromosome organization. Annals of Botany *82*, 57-66.

Narayan, R.K.J. (1982). Discontinuous DNA variation in the evolution of plant species: the genus *Lathyrus*. Evolution *36*, 877-891.

Nerkar, Y.S. (1972). Induced variation and response to selection for low neurotoxin content in *Lathyrus sativus*. Indian Journal of Genetics and Plant Breeding *32*, 175-180.

Nerkar, Y.S. (1973). Induced mutations of phylogenetic significance in *Lathyrus sativus*. Indian Journal of Genetics and Plant Breeding *33*, 324-325.

Nerkar, Y.S. (1976). Mutation studies in *Lathyrus sativus*. Indian Journal of Genetics and Plant Breeding *36*, 223-229.

Nerkar, Y.S. (1977a). Cytogenetical effects of gamma-rays, ethyl methane sulfonate and nitroso methyl urea in *Lathyrus sativus*. Indian Journal of Genetics and Plant Breeding *37*, 142-146.

Nerkar, Y.S. (1977b). Mutagenic effectiveness and efficiency of gamma-rays, ethyl methanesulfonate and nitroso methyl urea in *Lathyrus sativus*. Indian Journal of Genetics and Plant Breeding *37*, 137-141.

Noonan, S., and Savage, G. (1999). Oxalate content of foods and its effect on humans. Asia Pacific Journal of Clinical Nutrition *8*, 64-74.

Nunn, P.B., Lyddiard, J.R., and Christopher Perera, K.P. (2011). Brain glutathione as a target for aetiological factors in neurolathyrism and konzo. Food and Chemical Toxicology *49*, 662-667.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics *120*, 621-623.

Okuda, H., Lee, S.-D., Matsuura, Y., Zheng, Y., Sekiya, K., Takaku, T., Kameda, K., Hirose, K., Ohtani, K., Tanaka, O*., et al.* (1990). Biological activities of non-saponin compounds isolated from Korean red ginseng. Journal of Ginseng Research *14*, 157-161.

Ormandy, G.C., and Jope, R.S. (1990). Inhibition of phosphoinositide hydrolysis by the novel neurotoxin beta-N-oxalyl-L-alpha, beta-diaminopropionic acid (L-BOAA). Brain Research Bulletin *510*, 53-57.

Padmajaprasad, V., Kaladhar, M., and Bhat, R.V. (1997). Thermal isomerisation of β-*N*-oxalyl-L-α, β-diaminopropionic acid, the neurotoxin in *Lathyrus sativus*, during cooking. Food Chemistry *59*, 77-80.

Pai, K.S., and Ravindranath, V. (1993). L-BOAA induces selective inhibition of brain mitochondrial enzyme, NADH-dehydrogenase. Brain Research Bulletin *621*, 215-221.

Pandey, R., Chitale, M., Sharma, R., and Rastogi, N. (1995). Status of *Lathyrus* research in India. Paper presented at: Lathyrus Genetic Resources in Asia (Raipur, India: International Plant Genetic Resources Institute).

Parker, A.J., Mehta, T., Zarghami, N.S., Cusick, P.K., and Haskell, B.E. (1979). Acute neurotoxicity of the *Lathyrus sativus* neurotoxin, L-3-oxalylamino-2-aminopropionic acid, in the squirrel monkey. Toxicology and Applied Pharmacology *47*, 135-143.

Parry, M.A.J., Madgwick, P.J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A., Ouabbou, H*., et al.* (2009). Mutation discovery for crop improvement. Journal of Experimental Botany *60*, 2817-2825.

Pastor-Cavada, E., Juan, R., Pastor, J.E., Alaiz, M., and Vioque, J. (2011). Nutritional characteristics of seed proteins in 15 *Lathyrus* species (fabaceae) from Southern Spain. LWT - Food Science and Technology *44*, 1059-1064.

Pavan, S., Schiavulli, A., Appiano, M., Marcotrigiano, A.R., Cillo, F., Visser, R.G., Bai, Y., Lotti, C., and Ricciardi, L. (2011). Pea powdery mildew er1 resistance is associated to loss-of-function mutations at a MLO homologous locus. Theoretical and Applied Genetics *123*, 1425-1431.

Peña-Chocarro, L., Pérez Jordà, G., Morales Mateos, J., and Zapata, L. (2013). Neolithic plant use in the Western Mediterranean region: preliminary results from the AGRIWESTMED project. Annali di Botanica *3*, 135-141.

Perry, J.A., Wang, T.L., Welham, T.J., Gardner, S., Pike, J.M., Yoshida, S., and Parniske, M. (2003). A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiology *131*, 866-871.

Peyret, H., and Lomonossoff, G.P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Molecular Biology *83*, 51-58.

Piergiovanni, A.R., Lupo, F., and Zaccardelli, M. (2011). Environmental effect on yield, composition and technological seed traits of some Italian ecotypes of grass pea (*Lathyrus sativus* L.). Journal of the Science of Food and Agriculture *91*, 122-129.

Polignano, G., Bisignano, V., Tomaselli, V., Uggenti, P., Alba, V., and Della Gatta, C. (2009). Genotype × environment interaction in grass pea (*Lathyrus sativus* L.) lines. International Journal of Agronomy *2009*.

Polignano, G., Uggenti, P., Alba, V., Bisignano, V., and Della Gatta, C. (2005). Morpho-agronomic diversity in grasspea (*Lathyrus sativus* L.). Plant Genetic Resources: characterization and utilization *3*, 29-34.

Potrykus, I. (2010). Lessons from the 'Humanitarian Golden Rice' project: regulation prevents development of public good genetically engineered crop products. New biotechnology *27*, 466-472.

Pradesh, M. (2008). Non-uniform implementation of ban on *Lathyrus* cultivation in Indian states leading to unwarranted exposure to consumers. Current Science *94*, 570.

Pratap Rudra, M.P., Singh, M.R., Junaid, M.A., Jyothi, P., and Rao, S.L. (2004). Metabolism of dietary ODAP in humans may be responsible for the low incidence of neurolathyrism. Clinical Biochemistry *37*, 318-322.

Purchase, I.F.H. (2005). What determines the acceptability of genetically modified food that can improve human nutrition? Toxicology and Applied Pharmacology *207*, 19-27.

Quader, M., Singh, S., and Barat, G. (1987). Genetic analysis of BOAA content in *Lathyrus sativus* L. The Indian Journal of Genetics and Plant Breeding *47*, 275-279.

Quereshi, M.Y., Pilbeam, D.J., Evans, C.S., and Bell, E.A. (1977). The neurolathyrogen, α-amino-β-oxalylaminopropionic acid in legume seeds. Phytochemistry *16*, 477-479.

Ramachandran, S., Bairagi, A., and Ray, A.K. (2005). Improvement of nutritive value of grass pea (*Lathyrus sativus*) seed meal in the formulated diets for rohu, *Labeo rohita* (Hamilton) fingerlings after fermentation with a fish gut bacterium. Bioresource Technology *96*, 1465-1472.

Rao, S. (1978). A sensitive and specific colorimetric method for the determination of α, β-diaminopropionic acid and the *Lathyrus sativus* neurotoxin. Analytical Biochemistry *86*, 386-395.

Rao, S., Adiga, P., and Sarma, P. (1964). The isolation and characterization of β-N-oxalyl-L-α, β-diaminopropionic acid: a neurotoxin from the seeds of *Lathyrus sativus*. Biochemistry *3*, 432-436.

Rao, S.C., and Northup, B.K. (2011). Growth and nutritive value of grass pea in Oklahoma. Agronomy Journal *103*, 1692-1696.

Rao, S.L.N. (2011). A look at the brighter facets of beta-N-oxalyl-l-alpha,beta-diaminopropionic acid, homoarginine and the grass pea. Food and Chemical Toxicology *49*, 620-622.

Ray, D.K., Mueller, N.D., West, P.C., and Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. Plos One *8*, e66428.

Rietz, S., Bernsdorff, F.E.M., and Cai, D. (2012). Members of the germin-like protein family in Brassica napus are candidates for the initiation of an oxidative burst that impedes pathogenesis of *Sclerotinia sclerotiorum*. Journal of Experimental Botany *63*, 5507-5519.

Rispail, N., and Rubiales, D. (2016). Genome-wide identification and comparison of legume MLO gene family. Scientific Reports *6*, 32673.

Ross, S.M., Roy, D.N., and Spencer, P.S. (1985). beta-N-Oxalylamino-L-alanine: action on high-affinity transport of neurotransmitters in rat brain and spinal cord synaptosomes. Journal of Neurochemistry *44*, 886-892.

Ross, S.M., Roy, D.N., and Spencer, P.S. (1989). Beta-N-oxalylamino-L-alanine action on glutamate receptors. Journal of Neurochemistry *53*, 710-715.

Rowe, L.D., Ivie, G.W., DeLoach, J.R., and Foster, J.G. (1993). The toxic effects of mature flatpea (*Lathyrus sylvestris* L. cv. *Lathco*) on sheep. Veterinary and Human Toxicology *35*, 127-133.

Roy, D., and Bhat, R. (1975). Variation in neurotoxin, trypsin inhibitors and susceptibility to insect attack in varieties of *Lathyrus sativus* seeds. Environmental physiology & biochemistry *5*, 172.

Roy, D., and Rao, B.N. (1968). Distribution of alpha- and beta-isomers of N-oxalyl-alpha,beta-diamino propionic acid in some Indian varieties of *L. sativus*. Current Science *37*, 395-&.

Rybiński, W. (2003). Mutagenesis as a tool for improvement of traits in grasspea *(Lathyrus sativus* L.). Lathyrus lathyrism Newsletter, 27-31.

Rybinski, W., Blaszczak, W., and Fornal, J. (2006). Seed microstructure and genetic variation of characters in selected grass-pea mutants (*Lathyrus sativus* L.). International agrophysics *20*, 317.

Sainsbury, F., and Lomonossoff, G.P. (2008). Extremely high-level and rapid transient protein production in plants without the use of viral replication. Plant Physiology *148*, 1212-1218.

Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnology Journal *7*, 682-693.

Sakamoto, A., Nishimura, T., Miyaki, Y.-i., Watanabe, S., Takagi, H., Izumi, S., and Shimada, H. (2015). In vitro and in vivo evidence for oxalate oxidase activity of a germin-like protein from azalea. Biochemical and Biophysical Research Communications *458*, 536-542.

Santha, I.M., and Mehta, S.L. (2001). Development of low ODAP somaclones of *Lathyrus sativus*. Lathyrus lathyrism Newsletter, 42.

Sarker, A., El Moneim, A.A., and Maxted, N. (2001). Grasspea and Chicklings (Lathyrus L.). Book chapter in Plant Genetic Resources of Legumes in the Mediterranean, N. Maxted, and S.J. Bennett, eds. (Dordrecht: Springer Netherlands), pp. 159-180.

Sarwar, C., Malek, M., Sarker, A., and Hassan, M. (1995). Genetic resources of grasspea (*Lathyrus sativus* L.) in Bangladesh. Paper presented at: Lathyrus Genetic Resources in Asia (Raipur, India: International Plant Genetic Resources Institute).

Sastri, A. (2008). Varieties Developed and Releases by Indira Gandhi Krishi Vishwavidyalaya (Indira Gandhi Krishi Vishwavidyalya, Raipur).

Sawant, P., Jayade, V., and Patil, S. (2011). Line × tester analysis in *Lathyrus*. Journal of Food Legumes *24*, 41-45.

Scarth, R., and McVetty, P.B. (1999). Designer oil canola–a review of new food-grade *Brassica* oils with focus on high oleic, low linolenic types. Paper presented at: 10th International Rapeseed Congress (Canberra, Australia: Organising Committee of the 10th International Rapeseed Congress:).

Schenk, S.U., and Werner, D. (1991). β-(3-isoxazolin-5-on-2-yl)-alanine from *Pisum*: allelopathic properties and antimycotic bioassay. Phytochemistry *30*, 467-470.

Schultz, T.W., and Ranney, T.S. (1988). Structure-activity relationships for osteolathyrism: II. Effects of alkyl-substituted acid hydrazides. Toxicology *53*, 147-159.

Schulz, S., Keatinge, J.D.H., and Wells, G.J. (1999). Productivity and residual effects of legumes in rice-based cropping systems in a warm-temperate environment: I. Legume biomass production and N fixation. Field Crops Research *61*, 23-35.

Shah, T.M., Mirza, J.I., Haq, M.A., and Atta, B.M. (2008). Induced genetic variability in chickpea (*Cicer arietinum* L.). II. Comparative mutagenic effectiveness and efficiency of physical and chemical mutagens. Pakistan Journal of Botany *40*, 605-613.

Sharma, R., Chitale, M., Ganvir, G., Geda, A., and Pandey, R. (2000). Observations on the development of selection criterion for high yield and low neurotoxin in grass pea based on genetic resources. Lathyrus lathyrism Newsletter *1*, 15-16.

Shelton, J.M., Coleman, M.C., Herndon, N., Lu, N., Lam, E.T., Anantharaman, T., Sheth, P., and Brown, S.J. (2015). Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. BMC Genomics *16*, 1-16.

Sheridan, C. (2009). Doubts surround link between Bt cotton failure and farmer suicide. Nature Biotechnology *27*, 9-10.

Shiferaw, E. (2013). Development and cross-species amplification of grass pea EST-derived markers. African Crop Science Journal *21*, 153-160.

Shinomol, G.K., and Muralidhara (2007). Differential induction of oxidative impairments in brain regions of male mice following subchronic consumption of Khesari dhal (Lathyrus sativus) and detoxified Khesari dhal. Neurotoxicology *28*, 798-806.

Siddique, K.H.M., Hanbury, C.L., and Sarker, A. (2006). Registration of 'Ceora' Grass Pea Registration by CSSA. Crop Science *46*, 986-986.

Silvestre, S., de Sousa Araújo, S., Vaz Patto, M.C., and Marques da Silva, J. (2014). Performance index: an expeditious tool to screen for improved drought resistance in the *Lathyrus* genus. Journal of integrative plant biology *56*, 610-621.

Simons Jr, S.S., and Johnson, D.F. (1976). The structure of the fluorescent adduct formed in the reaction of o-phthalaldehyde and thiols with amines. Journal of the American Chemical Society *98*, 7098-7099.

Simpson, R. (2002). Mission in Afghanistan. The Medical Journal of Australia *177*, 633-637.

Singh, M., and Chaturvedi, S. (1987). Effectiveness and efficiency of mutagens alone or in combination with dimethyl sulphoxide in *Lathyrus sativus* Linn. Indian Journal of Agricultural Sciences *57*, 503-507.

Singh, O.V., Ghai, S., Paul, D., and Jain, R.K. (2006). Genetically modified crops: success, safety assessment, and public concern. Applied Microbiology and Biotechnology *71*, 598-607.

Singh, S.S., and Rao, S. (2013). Lessons from neurolathyrism: A disease of the past & the future of *Lathyrus sativus* (Khesari dal). The Indian journal of medical research *138*, 32.

Skiba, B., Gurung, A.M., and Pang, E.C. (2007). Genome mapping and molecular breeding in *Lathyrus*. Book chapter in Pulses, Sugar and Tuber Crops (Springer), pp. 123-132.

Smale, M., Byerlee, D., and Jayne, T. (2013). Maize revolutions in sub-Saharan Africa. Book chapter in an African green revolution (Springer), pp. 165-195.

Smartt, J. (1984). Evolution of grain legumes. I. Mediterranean pulses. Experimental Agriculture *20*, 275-296.

Smyth, S., and McHughen, A. (2008). Regulating innovative crop technologies in Canada: the case of regulating genetically modified crops. Plant Biotechnology Journal *6*, 213-225.

Sparvoli, F., Laureati, M., Pilu, R., Pagliarini, E., Toschi, I., Giuberti, G., Fortunati, P., Daminati, M.G., Cominelli, E., and Bollini, R. (2016). Exploitation of common bean flours

with low antinutrient content for making nutritionally enhanced biscuits. Frontiers in Plant Science *7*, 928.

Spencer, P.S., Ludolph, A., Dwivedi, M.P., Roy, D.N., Hugon, J., and Schaumburg, H.H. (1986). Lathyrism: Evidence for role of the neuroexcitatory amino acid BOAA. The Lancet *328*, 1066-1067.

Srivastava, R.P., and Srivastava, G.K. (2006). Accumulation of β-N-oxalyl amino alanine in *Lathyrus* during podding and maturation. Indian Journal of Agricultural Biochemistry *19*, 39-41.

Stokvis, E., Rosing, H., and Beijnen, J.H. (2005). Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not? Rapid communications in mass spectrometry *19*, 401-407.

Stone, G.D. (2007). The birth and death of traditional knowledge: Paradoxical effects of biotechnology in India. Book chapter in Biodiversity and the law: Intellectual property, biotechnology and traditional knowledge, C.R. McManis, ed. (Taylor & Francis Ltd.), pp. 207-238.

Sun, S., Fu, H., Wang, Z., Duan, C., Zong, X., and Zhu, Z. (2016a). Discovery of a novel er1 allele conferring powdery mildew resistance in Chinese pea (*Pisum sativum* L.) landraces. Plos One *11*, e0147624.

Sun, T., Xu, Y., Zhang, D., Zhuang, H., Wu, J., and Sun, G. (2016b). An acyltransferase gene that putatively functions in anthocyanin modification was horizontally transferred from Fabaceae into the genus *Cuscuta*. Plant Diversity *38*, 149-155.

Sun, X.-L., Yang, T., Guan, J.-P., Ma, Y., Jiang, J.-Y., Cao, R., Burlyaeva, M., Vishnyakova, M., Semenova, E., Bulyntsev, S.*, et al.* (2012). Development of 161 novel EST-SSR markers from *Lathyrus sativus* (Fabaceae). American Journal of Botany *99*, e379-e390.

Tadesse, W., and Bekele, E. (2003a). Phenotypic diversity of Ethiopian grass pea *(Lathyrus sativus* L.) in relation to geographical regions and altitudinal range. Genetic Resources and Crop Evolution *50*, 497-505.

Tadesse, W., and Bekele, E. (2003b). Variation and association of morphological and biochemical characters in grass pea (Lathyrus sativus L.). Euphytica *130*, 315-324.

Takos, A., Lai, D., Mikkelsen, L., Abou Hachem, M., Shelton, D., Motawia, M.S., Olsen, C.E., Wang, T.L., Martin, C., and Rook, F. (2010). Genetic screening identifies cyanogenesis-deficient mutants of *Lotus japonicus* and reveals enzymatic specificity in hydroxynitrile glucoside metabolism. Plant Cell *22*, 1605-1619.

Talukdar, D. (2009a). Dwarf mutations in grass pea (*Lathyrus sativus* L.): origin, morphology, inheritance and linkage studies. Journal of Genetics *88*, 165-175.

Talukdar, D. (2009b). Recent progress on genetic analysis of novel mutants and aneuploid research in grass pea (*Lathyrus sativus* L.). African Journal of Agricultural Research *14*, 1549-1559.

Talukdar, D. (2013). Cytogenetics of a reciprocal translocation integrating distichous pedicel and tendril-less leaf mutations in *Lathyrus sativus* L. Caryologia *66*, 21-30.

Tan, X., Yan, S., Tan, R., Zhang, Z., Wang, Z., and Chen, J. (2014). Characterization and expression of a GDSL-like lipase gene from *Brassica napus* in *Nicotiana benthamiana*. The protein journal *33*, 18-23.

Tavoletti, S., Iommarini, L., Crino, P., and Granati, E. (2005). Collection and evaluation of grasspea (Lathyrus sativus L.) germplasm of central Italy. Plant Breeding *124*, 388-391.

Tay, J., Valenzuela, A., and Venegas, F. (2000). Collecting and evaluating Chilean germplasm of grasspea (*Lathyrus sativus* L.). Lathyrus lathyrism Newsletter, 21.

Thirtle, C., Beyers, L., Ismael, Y., and Piesse, J. (2003). Can GM-technologies help the poor? The impact of Bt cotton in Makhathini Flats, KwaZulu-Natal. World Development *31*, 717-732.

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S., and Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biology *7*, 19-19.

Till, B.J., Zerr, T., Comai, L., and Henikoff, S. (2006). A protocol for TILLING and Ecotilling in plants and animals. Nat Protoc *1*, 2465-2477.

Tilman, D., Balzer, C., Hill, J., and Befort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences *108*, 20260-20264.

Tiwari, K., and Campbell, C. (1996). Inheritance of neurotoxin (ODAP) content, flower and seed coat colour in grass pea (*Lathyrus sativus* L.). Euphytica *91*, 195-203.

Tripathy, S.K., Ranjan, R., and Lenka, D. (2012). Effectiveness and efficiency of single and combined treatments of physical and chemical mutagens in grasspea (*Lathyrus sativus* L.). World Applied Sciences Journal *20*, 738-741.

Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K.J., Lieberman, M., Fass, J., Uauy, C., Tran, R.K*., et al.* (2011). Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiology *156*, 1257-1268.

Tsegaye, D., Tadesse, W., and Bayable, M. (2005). Performance of grass pea (*Lathyrus sativus* L.) somaclones at Adet, northwest Ethiopia. Lathyrus lathyrism Newsletter *4*, 5-6.

Tshala-Katumbay, D., Mwanza, J.-C., Rohlman, D.S., Maestre, G., and Oriá, R.B. (2015). A global perspective on the influence of environmental exposures on the nervous system. Nature *527*, S187-S192.

Tshala-Katumbay, D., and Spencer, P.S. (2007). Toxic disorders of the upper motor neuron system. Handbook of clinical neurology *82*, 353-372.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. Frontiers in Physiology *3*, 347.

Tuominen, L.K., Johnson, V.E., and Tsai, C.-J. (2011). Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. BMC Genomics *12*, 1-17.

Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M., and Nakayama, T. (2007). Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis. Journal of Biological Chemistry *282*.

Valamoti, S.M., Moniaki, K., and Karthanou, A. (2010). Processing and consumption of pulses in prehistoric Greece: archaeobotanical, experimental and ethnographic evidence. Paper presented at: 15th Conference of the International Work Group for Palaeoethnobotany, Wilhemshaven, Germany.

Van Moorhem, M., Lambein, F., and Leybaert, L. (2011). Unraveling the mechanism of beta-N-oxalyl-alpha,beta-diaminopropionic acid (beta-ODAP) induced excitotoxicity and oxidative stress, relevance for neurolathyrism prevention. Food and Chemical Toxicology *49*, 550-555.

Van Zeist, W., and Bakker-Heeres, J. (1985). Archaeobotanical studies in the Levant. 4. Bronze Age sites on the north Syrian Euphrates. Palaeohistoria *27*, 247-316.

Vavilov, N.I. (1927). Geographical regularities in the distribution of the genes of cultivated plants. Bulletin of Applied Botany, of Genetics and Plant-Breeding *17*, 411-428.

Vaz Patto, M.C., Fernández‑Aparicio, M., Moral, A., and Rubiales, D. (2006). Characterization of resistance to powdery mildew (*Erysiphe pisi*) in a germplasm collection of *Lathyrus sativus*. Plant Breeding *125*, 308-310.

Vaz Patto, M.C., and Rubiales, D. (2014). *Lathyrus* diversity: available resources with relevance to crop improvement – *L. sativus* and *L. cicera* as case studies. Annals of Botany *113*, 895-908.

Verdier, J., Torres‑Jerez, I., Wang, M., Andriankaja, A., Allen, S.N., He, J., Tang, Y., Murray, J.D., and Udvardi, M.K. (2013). Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. The Plant Journal *74*, 351-362.

Vreugdenhil, D., Vreugdenhil, A., and Tilahun, T. (2012). Ecosystems map of Ethiopia. In Gap Analysis of the Protected Areas System of Ethiopia (Addis Ababa: World Institute for Conservation and Environment), pp. 1-33.

Waghmare, V., and Mehra, R. (2001). Induced chlorophyll mutations, mutagenic effectiveness and efficiency in *Lathyrus sativus* L. The Indian Journal of Genetics and Plant Breeding *61*, 53-56.

Wale, M., and Gedif, A. (2013). *Acyrthosiphon pisum* (Harris)(Homoptera: Aphididae) infestation level and damage on grass pea, *Lathyrus sativus* L., in West Gojam, Ethiopia. Ethiopian Journal of Science and Technology *6*, 69-78.

Walker, K., Long, R., and Croteau, R. (2002). The final acylation step in Taxol biosynthesis: Cloning of the taxoid C13-side-chain N-benzoyltransferase from Taxus. Proceedings of the National Academy of Sciences *99*, 9166-9171.

Waltz, E. (2016). Gene-edited CRISPR mushroom escapes US regulation. In Nature News (Nature), pp. 293.

Wang, F., Yang, T., Burlyaeva, M., Li, L., Jiang, J., Fang, L., Redden, R., and Zong, X. (2015). Genetic diversity of grasspea and its relative species revealed by SSR Markers. Plos One *10*, e0118542.

Wang, T.L., Uauy, C., Robson, F., and Till, B. (2012). TILLING in extremis. Plant Biotechnology Journal *10*, 761-772.

Wang, X., Warkentin, T.D., Briggs, C.J., Oomah, B.D., Campbell, C.G., and Woods, S. (1998). Trypsin inhibitor activity in field pea (*Pisum sativum* L.) and grass pea (*Lathyrus sativus* L.). Journal of Agricultural and Food Chemistry *46*, 2620-2623.

Wang, Z., Yang, J.Y., and Chen, Q. (2014). Development of Dencichine Plus Yunnan Baiyao band-aid. Paper presented at: Advanced Materials Research (Trans Tech Publ).

Weeden, N.F. (2007). Genetic Changes Accompanying the Domestication of Pisum sativum: Is there a Common Genetic Basis to the 'Domestication Syndrome' for Legumes? Annals of Botany *100*, 1017-1025.

Weisenfeld, N.I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., Sogoloff, B., Tabbaa, D., Williams, L., Russ, C*., et al.* (2014). Comprehensive variation discovery in single human genomes. Nature Genetics *46*, 1350-1355.

Weiss, J.H., Koh, J.Y., and Choi, D.W. (1989). Neurotoxicity of beta-N-methylamino-L-alanine (BMAA) and beta-N-oxalylamino-L-alanine (BOAA) on cultured cortical neurons. Brain Research *497*, 64-71.

Wesseler, J., and Zilberman, D. (2016). Golden Rice: no progress to be seen. Do we still need it? Environment and Development Economics *22*, 107-109.

Westmore, A., and Weisz, G.M. (2013). Medical research undertaken in captivity: a form of resistance to imprisonment and attempted extermination. War & Society *28*, 89-112.

Williams, B., Kabbage, M., Kim, H.-J., Britt, R., and Dickman, M.B. (2011). Tipping the balance: *Sclerotinia sclerotiorum* secreted oxalic acid suppresses host defenses by manipulating the host redox environment. PLoS Pathogens *7*, e1002107.

Williams, S., Vivero Pol, J., Spawls, S., Shimelis, A., Kelbessa, E., Mittermeier, R., Robles Gil, P., Hoffmann, M., Pilgrim, J., and Brooks, T. (2004). Ethiopian highlands. Book chapter in Hotspots Revisited: Earths Biologically Richest and Most Endangered Ecoregions, R.A. Mittermeier, P.R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C.G. Mittermeier, J. Lamoreux, and G.A. Da Fonseca, eds. (Mexico City: Conservation International).

Willimott, S. (1933). An investigation of solanine poisoning. Analyst *58*, 431-439.

Wilson, W.M., and Dufour, D.L. (2002). Why "Bitter" Cassava? Productivity of "Bitter" and "Sweet" Cassava in a Tukanoan Indian Settlement in the Northwest Amazon. Economic Botany *56*, 49-57.

Woldeamanuel, Y.W., Hassan, A., and Zenebe, G. (2012). Neurolathyrism: two Ethiopian case reports and review of the literature. Journal of neurology *259*, 1263-1268.

Wolt, J.D., Wang, K., and Yang, B. (2016). The regulatory status of genome-edited crops. Plant Biotechnology Journal *14*, 510-518.

Wu, G., Bowlus, S.B., Kyung, S.K., and Haskell, B.E. (1976). L-2-oxalylamino-3-aminopropionic acid, an isomer of *Lathyrus sativus* neurotoxin. Phytochemistry *15*, 1257-1259.

Xie, G.X., Qiu, Y.P., Qiu, M.F., Gao, X.F., Liu, Y.M., and Jia, W. (2007). Analysis of dencichine in *Panax notoginseng* by gas chromatography-mass spectrometry with ethyl chloroformate derivatization. Journal of Pharmaceutical and Biomedical Analysis *43*, 920-925.

Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR–Cas system. Molecular Plant *6*, 1975-1983.

Xing, G., Cui, K., Ji, L., Wang, Y., and Li, Z. (2001). Water stress and accumulation of β-N-oxalyl-l-α,β-diaminopropionic acid in grass pea (*Lathyrus sativus*). Journal of Agricultural and Food Chemistry *49*, 216-220.

Xiong, J.L., Bai, X., Batool, A., Kong, H.Y., Tan, R.Y., Wang, Y.F., Li, Z.X., and Xiong, Y.C. (2014). [Ecological function and application of toxin beta-ODAP in grass pea (*Lathyrus sativus*)]. Ying Yong Sheng Tai Xue Bao *25*, 1197-1205.

Xiong, J.L., Xiong, Y.C., Bai, X., Kong, H.Y., Tan, R.Y., Zhu, H., Siddique, K.H., Wang, J.Y., and Turner, N.C. (2015). Genotypic variation in the concentration of beta-N-oxalyl-L-alpha,beta-diaminopropionic acid (beta-ODAP) in grass pea (*Lathyrus sativus* L.) seeds is associated with an accumulation of leaf and pod beta-ODAP during vegetative and reproductive stages at three levels of water stress. Journal of Agricultural and Food Chemistry *63*, 6133-6141.

Xiong, Y.C., Xing, G.M., Li, F.M., Wang, S.M., Fan, X.W., Li, Z.X., and Wang, Y.F. (2006). Abscisic acid promotes accumulation of toxin ODAP in relation to free spermine level in grass pea seedlings (*Lathyrus sativus* L.). Plant Physiology and Biochemistry *44*, 161-169.

Yadav, V.K., and Mchta, S.L. (1995). *Lathyrus sativus*: A future pulse crop free of neurotoxin. Current Science *68*, 288-292.

Yan, Z., Wang, Y., Jiao, C., Li, F., Liang, Y., and Li, Z. (2005). High-performance liquid chromatographic analysis of neurotoxin β-N-oxalyl-α, β-diaminopropionic acid (β-ODAP), its non-neurotoxic isomer α-ODAP and other free amino acids in *Lathyrus sativus*. Chromatographia *61*, 231-236.

Yan, Z.Y., Spencer, P.S., Li, Z.X., Liang, Y.M., Wang, Y.F., Wang, C.Y., and Li, F.M. (2006). *Lathyrus sativus* (grass pea) and its neurotoxin ODAP. Phytochemistry *67*, 107-121.

Yang, H.-M., and Zhang, X.-Y. (2005). Considerations on the reintroduction of grass pea in China. Lathyrus lathyrism Newsletter *4*, 22-26.

Yang, Q., Reinhard, K., Schiltz, E., and Matern, U. (1997). Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. Plant Molecular Biology *35*, 777-789.

Yang, T., Jiang, J.Y., Burlyaeva, M., Hu, J.G., Coyne, C.J., Kumar, S., Redden, R., Sun, X.L., Wang, F., Chang, J.W.*, et al.* (2014). Large-scale microsatellite development in grasspea (*Lathyrus sativus* L.), an orphan legume of the arid areas. BMC Plant Biology *14*.

Yigzaw, Y., Gorton, L., Solomon, T., and Akalu, G. (2004). Fermentation of seeds of Teff (*Eragrostis teff*), grass-pea (*Lathyrus sativus*), and their mixtures: aspects of nutrition and food safety. Journal of Agricultural and Food Chemistry *52*, 1163-1169.

Yu, L., Jiang, J., Zhang, C., Jiang, L., Ye, N., Lu, Y., Yang, G., Liu, E., Peng, C., He, Z.*, et al.* (2010). Glyoxylate rather than ascorbate is an efficient precursor for oxalate biosynthesis in rice. Journal of Experimental Botany *61*, 1625-1634.

Yu, X.-H., Chen, M.-H., and Liu, C.-J. (2008). Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in *Medicago truncatula*. The Plant Journal *55*, 382-396.

Zhelyazkova, T., Pavlov, D., Delchev, G., and Stoyanova, A. (2016). Productivity and yield stability of six grain legumes in the moderateclimatic conditions in Bulgaria. Scientific Papers Series A Agronomy *LIX*, 478-487.

Zhu, J., Liu, S.K., Fu, C.M., and Li, Z.W. (2006). [Analysis of dencichine by HPLC with pre-column derivatization]. Zhongguo Zhong Yao Za Zhi *31*, 1865-1868.